MIDAS TOUCH

Doping with gold modulates the electronic properties
of a semiconductor nanocrystal superlattice   **PAGES 418 & 450**

3 5 ▷

9 770028 083095

FDA vulnerability revealed

A politically charged advisory committee meeting may have tipped the scales in favour of a mildly effective female libido drug.

Loud applause erupted when an advisory panel recommended in early June that the US Food and Drug Administration (FDA) approve the first medication to treat low libido in women. Many more must have cheered when the FDA followed the committee's advice and approved the prescription drug, which will go by the name Addyi, on 18 August.

The decision comes as the FDA faces greater pressure to approve more drugs, especially for conditions that lack other treatments. The agency has accepted 89% of applications for new drug approvals so far this year compared with a rate of 50% in 2008, according to an analysis commissioned by Forbes. Legislation in the US Congress — the 21st Century Cures Act — could encourage even more approvals. The FDA is also receiving more input from patients to gather perspectives on their diseases and available treatments. In October 2014, for instance, it convened a two-day public meeting on female sexual dysfunction, which Addyi (flibanserin) aims to address.

But critics say that previous accusations that the FDA is biased against women had more influence on the drug's approval than did the data on its safety and efficacy. They blame a coalition of non-profit organizations and drug companies called Even The Score — which includes Sprout Pharmaceuticals, the maker of Addyi — for leading a campaign that made approval an issue of gender equality. One of the group's slogans highlighted a count of the number of drugs available to treat sexual problems in men versus women: 26 to 0.

Addyi was found to be only modestly effective in improving female sexual desire during clinical trials. On average, women taking it reported having from 0.5 to 1 additional satisfying sexual experiences per month, compared with a group taking a placebo. Participants also reported slightly higher levels of desire and slightly lower levels of distress. Small as such benefits may be, they should not be dismissed.

At the June meeting, the FDA rejected the idea that it had a gender bias. It also warned that rare sudden drops in blood pressure and a poorly understood interaction with alcohol are outstanding concerns for Addyi. Incredibly, Sprout's alcohol-interaction study included 23 men and only 2 women. As part of its approval, the FDA is requiring Sprout — which Canadian drug maker Valeant bought for US$1 billion two days after Addyi's approval — to run three more studies on alcohol interaction after the drug goes on sale.

The vote of approval came after impassioned testimony from women who had taken the drug in trials, physicians who could offer no good alternatives to patients, and women who had experienced low libido for decades. The drug's supporters pointed to an unmet need for treatment options and said that it is only fair to have a choice about whether or not to take the drug. These points seemed to resonate with the advisory committee, which intimated reluctant support: one member who voted 'yes' noted that if Addyi were "the seventh drug in the class, it would be a very different discussion".

Many complain that the FDA acts either too fast or too slow. With ever more ways for new drugs to be approved, the agency will be expected to make more decisions more quickly. Addyi's approval reflects and encourages a trend of increased pressures from targeted campaigns. The agency and its committees will need resilience to ensure that such pressures do not overwhelm evidence-based caution. ∎

Heroism in Syria

A tribute to scholars of extraordinary courage.

During the Second World War, a group of historians, curators and professors risked life and limb to protect Europe's art and culture from Nazi plunderers. It is because of these 'monuments men' and women that future generations can appreciate works such as Michelangelo's sculpture *Madonna of Bruges*, Johannes Vermeer's painting *The Astronomer* and countless others.

The murder of the Syrian archaeologist Khaled al-Asaad last week is a brutal reminder of the risks that professionals have taken to protect the world's cultural heritage. Al-Asaad was known as 'Mr Palmyra' for his more than 40 years of work at that Roman Empire site in Syria. It is now occupied by Islamist militants, who last Sunday destroyed one of its temples. Al-Asaad was killed in part because he would not reveal the location of hidden artefacts, according to news reports. A week before, museum official Qasem Abdullah Yehiya was killed in a rocket attack on the Damascus Citadel and the National Museum of Damascus.

Across Syria and Iraq, a new generation of cultural guardians are risking their lives to safeguard museums, archaeological and other cultural sites from destruction and to document the damage. Their task is enormous. The destruction by Islamist terrorist group ISIS of sites such as the ancient Assyrian city of Nimrud has commanded most headlines, but its embrace of organized looting is just as great a threat. The collapse of civil society in Syria and Iraq has left hundreds of archaeological sites unguarded and susceptible to looting, encroachment and neglect.

Most of the protection efforts are necessarily secretive, but the stories that have emerged are truly heroic. Earlier this year, a group of Syrian professionals and volunteers secured a collection of Roman and Byzantine mosaics from the third to sixth centuries AD at the Ma'arra Mosaic Museum, near Aleppo, Syria. In February, *The Wall Street Journal* profiled a secretive band of Syrian academics who catalogue archaeological sites and track looted artefacts, sometimes donning disguises.

All of us who find inspiration in the cultural history of humanity, now and in the future, owe a huge debt to the sacrifices of these people. ∎

We must build resilience into our communities

Innovative approaches can better equip society to deal with natural disasters and other shocks, says **Erwann Michel–Kerjan**.

As part of a series of initiatives marking the tenth anniversary of Hurricane Katrina, this week I will join the mayor of New Orleans and other business and civic leaders in Louisiana to discuss publicly how science can help to prevent such devastation from happening again.

Katrina remains the most destructive natural disaster in recent US history — with more than 1,800 deaths, US$125 billion in economic losses and enduring social and political impacts. When the levee system failed, water from the Mississippi River, pushed by the strong winds and storm surge from Katrina, flooded a large portion of the city. Because many dwellings behind the barrier were built below sea level, it took days for the water to recede, causing permanent destruction. A then-ill-prepared government was slow to react, adding to the tragedy. A year later, New Orleans had lost half of its population, who moved to other cities — and it has yet to be totally restored.

Three questions are central. What have we learnt from Katrina? What new scientific knowledge can help us to be better prepared? And how do we encourage investment to make our communities more resilient?

One lesson was taken on board very quickly — that it is not enough for researchers to predict extreme weather events. They must also better communicate about risks, and plans must be in place to act on these risks. Although comparisons are hard to make, the contrast with Hurricane Sandy, which struck New York in 2012, is clear. Just 117 people in the United States died in Sandy, after the state and federal government (and the media) took the warnings from forecasters seriously and pre-positioned emergency response.

The difference in crisis management between 2005 and 2012 largely came down to improved governance and political leadership — lessons learnt in the seven years between the two events, as well as from other natural disasters, pandemics, oil spills and nuclear meltdowns around the world. No head of state wants to be remembered for another Katrina.

Predicting catastrophes is not an easy task, but we are getting better at it. This is good news, because in today's world, waiting to know for sure is a luxury that decision makers cannot afford. Some uncertainty surrounding the nature, likelihood and impact of adverse events will always remain, but it cannot be taken as an excuse for inaction.

New storm models better predict wind speed at points in space; storm-surge models are also improving. Climate predictions increasingly focus on regional impacts, which are much more relevant to local conditions than are global models. Social science, too, has made huge progress in understanding individual and collective behaviours when faced with disaster risks. Yet we must strive to connect these dots.

There are also major knowledge gaps that

> NO HEAD OF STATE WANTS TO BE **REMEMBERED** FOR ANOTHER **KATRINA.**

science can address. Take flooding, which has affected more people than any other natural hazard. Modernizing our flood maps is one massive but essential undertaking currently under way; an exposure analysis of all buildings, critical infrastructure and other assets in flood-prone areas is also key. It requires granular (hence costly) elevation data for each asset, of the kind typically gathered by a surveyor on site. Clearly, this is not economically practical on a large scale. But new remote-sensing techniques, such as laser and radar-based measurements collected by aircraft, drones or satellites, makes it feasible. And as this technology continues to develop, it will become cheaper.

My third point is on encouraging good investment. Perhaps the greatest contribution that research can make is to improve a region's resilience — and here, an apt mantra holds: what gets measured gets managed. Resilience must be captured holistically to spur financial interest and innovation.

A 'five capital' — 5 C — metric is essential. Physical capital includes the indirect products of economic activity, such as infrastructure; financial capital assesses financial protection and diversity of income sources; human capital pertains to the education, skills and health of people; social capital accounts for social relationships, leadership and governance structure; and natural capital includes land productivity, water and biological resources and actions to sustain them.

Research across many disciplines can gather data on the 5 Cs, to provide a baseline, measure progress and test policies. An agreed-upon set of metrics will enable a specific community to be given a resilience score that can be compared with those of others and tracked.

Our team at Wharton is piloting this approach as part of a global flood-resilience programme in Mexico, Peru, Indonesia and Nepal supported by the insurer Zurich, working with the Red Cross, the International Institute for Applied Systems Analysis and the non-governmental organization Practical Action. The US National Academy of Sciences is helping us to extend the idea to the United States through its Resilient America Roundtable.

In collaboration with the Rockefeller Foundation in New York, a growing number of cities, including New Orleans, are appointing chief resilience officers — who have similar roles to chief risk officers in companies — and the 5 C measurement tool should help them.

New Orleans native and jazz legend Louis Armstrong once said: "The memory of things gone is important to a jazz musician."

This week in Louisiana, it is important to all of us. ■

Erwann Michel-Kerjan *is executive director of the Center for Risk Management and Decision Processes at the Wharton School of the University of Pennsylvania in Philadelphia, USA, and co-author (with H. Kunreuther) of* At War with the Weather (MIT Press). *e-mail: erwannmk@wharton.upenn.edu*

CHEMISTRY

Better catalyst for carbon conversion

A porous, crystalline compound can speed up the conversion of carbon dioxide to carbon monoxide in water.

Omar Yaghi and Christopher Chang at the University of California, Berkeley, and their colleagues used structures called covalent organic frameworks (COFs) — grid-like arrangements of carbon, nitrogen and other light elements — to turn CO_2 into carbon monoxide. COFs containing cobalt resulted in conversion activity that was 60 times higher than that of a molecular cobalt complex that did not contain COFs.

Chemically converting CO_2 into a useful carbon product is a goal for clean energy. Catalysts built using COFs could be tuned further to boost other reactions, the authors say.
Science http://doi.org/6zr (2015)

ANIMAL BEHAVIOUR

Stinging cells help jellyfish to mate

Some box jellyfish display elaborate mating behaviours and even use their toxic stinging cells to ensure successful fertilization.

Many jellyfish reproduce using external fertilization, but in a few box jellyfish, fertilization can occur internally. In one species (*Copula sivickisi*; **pictured**), the male transfers a sperm

package into the female's stomach to fertilize the eggs after the animals entangle their tentacles. The females then lay strands of embryos.

Anders Garm and his team at the University of Copenhagen studied the sex organs of the jellyfish under a microscope and located the stinging cells in the sperm package and in the female gonad. The sperm package becomes attached to the female gonad and sperm cells are partly digested, releasing their nuclei, which are then taken up by the female sex organ. The stinging cells probably help the sperm package to attach, and protect the embryos once they are laid.
J. Morphol. 276, **1055–1064 (2015)**

ANIMAL BEHAVIOUR

Hummingbirds sip using mini pumps

Hummingbirds draw nectar into their bills using long tongues that act like tiny pumps.

It was long thought that liquid travels passively up the birds' tongues without suction. But Alejandro Rico-Guevara and his colleagues at the University of Connecticut in Storrs found a different mechanism when they captured slow-motion films of 18 species of hummingbirds feeding from transparent artificial flowers

(pictured is *Anthracothorax nigricollis*). The videos showed that two long grooves in the birds' tongues are squashed closed by the tip of the bill as the tongue protrudes. When dipped into the nectar, the grooves spring back open, creating a force that draws the liquid in.

This pumping technique allows the animals to sip nectar much faster than the passive action.
Proc. R. Soc. B 282, **20151014 (2015)**

ASTROPHYSICS

Dark-energy search narrows

Two groups have tightened the limits on the search for elusive dark matter and dark energy, the mysterious force accelerating the expansion of the Universe.

Physicists have proposed that dark energy could come from a 'chameleon' field: a force that would act in the low density of space but would not be felt in the relatively dense environment of the lab. Holger Müller at the University of California, Berkeley, and his team used a laser to measure the forces between

caesium atoms and a sphere of aluminium in an ultrahigh vacuum — a set-up designed to mimic outer space. They found no evidence of new forces, ruling out many chameleon fields that would explain the Universe's acceleration.

A second group, working on the Xenon100 experiment at the Gran Sasso National Laboratory, Italy, found no evidence for dark matter interacting with atoms inside its underground tank of liquid xenon. The group's findings rule out three kinds of dark matter as the source of a signal seen previously in another experiment at the same lab.
Science 349, **849–851; 851–854 (2015)**

HUMAN EVOLUTION

Old finger with modern traits

A 1.84-million-year-old finger bone from Tanzania is the oldest known hominin hand bone with human-like features.

Ancient human relatives used stone tools 2 million to 3 million years ago, but had hands that were suited to living in trees. A team led by Manuel Domínguez-Rodrigo at Madrid's Institute of Evolution in Africa found a finger bone at a site in Olduvai Gorge, Tanzania, in 2013. The bone, named OH 86, is longer and straighter than those of earlier australopiths and similarly aged *Homo habilis* hand bones found nearby. Its closest match in size and shape in the fossil record is a finger bone from early *Homo sapiens*.

The finding suggests that some human-like traits emerged early in human evolution, and that a modern-looking hominin lived alongside more primitive-bodied creatures in East Africa some 2 million years ago.
Nature Commun. 6, **7987** (2015)

ATMOSPHERIC SCIENCE

Carbon dioxide levels peak up high

The carbon dioxide concentration in Earth's upper atmosphere is increasing at more than twice the average rate observed at the surface.

Jia Yue of Hampton University in Virginia and his colleagues analysed CO_2 measurements at different atmospheric heights and latitudes between 2002 and 2014 using a satellite-borne infrared radiometer. They found that CO_2 levels increased at a peak rate of 12% per decade at a height of 110 kilometres, with more pronounced rises in the Northern Hemisphere than in the Southern Hemisphere. The rate of CO_2 increase in the atmosphere below 80 kilometres is only about 5% per decade.

The fast build-up of CO_2 in the upper atmosphere, which climate models fail to reproduce, points to increased vertical mixing. This will probably not affect greenhouse warming at Earth's surface, but could eventually change the upper atmosphere's density and alter the altitude of satellites in orbit, the authors suggest.
Geophys. Res. Lett. http://doi.org/6x5 (2015)

MEDICAL MICROBIOLOGY

Lung pathogen evolves in isolation

Bacteria that infect the lungs of people with cystic fibrosis evolve into different forms in various parts of the lungs.

Pradeep Singh at the University of Washington in Seattle and his team dissected the infected lungs of ten people with the disease who were having lung transplants, and analysed the genomes of *Pseudomonas aeruginosa* bacteria sampled from different areas of the lung. Comparing areas that were mildly and severely diseased, the team found distinct lineages of the bacterium with variations in virulence, antibiotic resistance and other factors. They conclude that this diversity arose because bacterial cells became isolated in lung regions where they then evolved separately.

This regional diversification could enhance resistance to drugs and human defences, the authors say.
Cell Host Microbe http://doi.org/6zw (2015)

INFORMATION TECHNOLOGY

Suspended rods serve as bits

Rod-shaped nanoparticles suspended in water can store the zeroes and ones of digital computing on the basis of the rods' physical location.

Most digital memories are made of solid matter. But Madhavi Krishnan at the University of Zurich in Switzerland and her colleagues stored bits of information using the position and 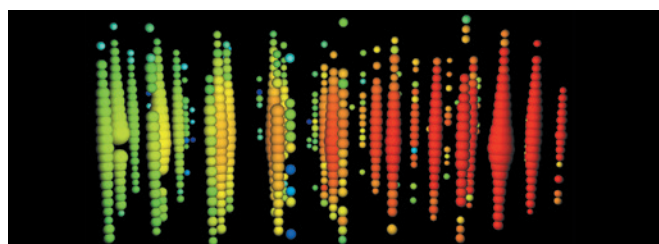orientation of small silver rods suspended in solution between plates that were 150 nanometres apart. The rods levitate at the mouth of one of two holes in the plates that form a T shape. A short electrical pulse or a beam of light can move a nanoparticle from one hole to the other.

By adjusting the geometry of the T, the team could control the lifetime of the stored bit, which they calculate could be as much as 30 years for their arrangement.
Nature Nanotech. http://doi.org/6x2 (2015)

ASTROPHYSICS

Cosmic neutrinos abound

Super-high-energy neutrinos from outside the Milky Way pepper Earth from all directions.

Neutrinos are created in the Universe's most violent environments and travel through it almost unimpeded, providing a way to study distant astronomical objects. A team at the IceCube Neutrino Observatory at the South Pole reported the first evidence for neutrinos from outside the Galaxy in 2013, using observations of the Southern Hemisphere sky. Now the team has used a different method to detect more than ten super-high-energy neutrinos from the Northern Hemisphere. IceCube detects tracks of light (**pictured**) when neutrinos interact with particles in the ice that forms part of the detector.

As with the previous result, the neutrinos come from all over the sky rather than being concentrated in the Galactic plane, bolstering the claim that they come from farther afield, say the authors.
Phys. Rev. Lett. 115, **081102** (2015)

➲ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

POLICY

Methane curbs

The US Environmental Protection Agency on 18 August proposed regulations to reduce emissions of methane and smog-forming volatile organic compounds (VOCs) from the oil and gas industry. The regulations would be the first to directly restrict methane emissions, and build on a 2012 rule that sought to curb VOCs from hydraulic fracturing (fracking) to extract natural gas. Combined, the two regulations could reduce the oil and gas sector's methane emissions by up to 30% by 2025, compared with 2012 levels. They are part of a larger effort by the White House to reduce national methane emissions by 40–45% by 2025. See go.nature.com/o6uzlj for more.

Solar power boost

US President Barack Obama has announced plans to promote and develop solar technology. The plans, laid out at the National Clean Energy Summit in Las Vegas, Nevada, on 24 August, include making it easier for homeowners to borrow money for energy upgrades, and US$24 million to fund research seeking to improve the efficiency of solar power. Obama criticized opponents of his energy polices for "wanting to protect an outdated status quo". The announcement comes on the heels of the Clean Power Plan, which aims for 20% of US electricity generation to come from renewables — excluding hydropower — by 2030.

Stem-cell guidelines

China's National Health and Family Planning Commission delivered long-awaited guidelines on the use of stem cells in China's clinics on 21 August. State media say that the measures will rein in the "wild" use of stem cells. The guidelines outline requirements for small-scale clinical studies using stem cells, including the use of clinical-grade cells, and forbid charging people in the study for the treatment. The studies will be restricted to authorized hospitals. Some researchers are embracing the measures as they offer an easy path to clinical studies of stem cells; previously, it was unclear what constituted a legal study. But scientists also say that the guidelines lack enforceable rules to prevent rogue clinics from using unapproved stem-cell therapies.

Female libido drug

The US Food and Drug Administration (FDA) approved the first drug for boosting female sexual desire on 18 August. The drug, Sprout Pharmaceuticals' flibanserin, acts on serotonin receptors in the brain and will be marketed as Addyi. The approval process has been arduous and controversial — many experts have questioned flibanserin's efficacy even as advocacy organizations pushed for the drug's approval. The FDA had twice rejected flibanserin because of safety and efficacy concerns. On 20 August, Canadian drug firm Valeant bought Sprout, of Raleigh, North Carolina, for US$1 billion. See page 387 for more.

Supplies soar to space

The six astronauts aboard the International Space Station (ISS) received much-needed supplies on 24 August, when the Japan Aerospace Exploration Agency docked its uncrewed ship Kounotori 5 to the ISS orbiting outpost. Kounotori 5 ('white stork' in Japanese) launched from Japan's Tanegashima Space Center on 19 August, carrying food, water and other provisions for the ISS astronauts, who were running short after three high-profile rocket failures in the past year. Also on the cargo carrier were a Japanese cosmic-ray telescope, a Brazilian nanosatellite and 14 Earth-imaging satellites designed by Planet Labs in San Francisco, California.

RESEARCH

Drifting science

An experiment that set two researchers adrift on an ice floe in the Arctic for a year ended in Norway's Svalbard archipelago on 22 August. Geophysicist Yngve Kristofferson of the University of Bergen, Norway, and Audun Tholfsen, a scientist in Longyearbyen, Svalbard, had drifted 2,200 kilometres from their starting point above the underwater Lomonosov Ridge, gathering seismic-reflection measurements and core samples to probe the ridge's geology. The pair had a hovercraft to gather data and rebuild their camp, which was twice destroyed by shifting ice. The ice-drift station, known as FRAM, also collected oceanographic and atmospheric measurements along the way.

PEOPLE

'Mr Palmyra' killed

Khaled al-Asaad, the Syrian archaeologist in charge of the ancient Roman site of Palmyra, was executed by Islamist militants on 18 August. Al-Asaad, who had worked at the site for more than 40 years, had been held captive for about a month and was killed, in part, because he would not reveal the location of hidden antiquities,

according to news reports. On 23 August, Islamist terrorist group ISIS blew up the almost 2,000-year-old Baalshamin temple in Palmyra. Irina Bokova, director-general of the United Nations Educational, Scientific and Cultural Organization, deplored the killing and described the temple's destruction as "a new war crime and an immense loss for the Syrian people and for humanity". See page 387 for more.

EVENTS

MERS resurgence

Saudi Arabia is facing a significant outbreak of Middle East respiratory syndrome (MERS), as an estimated 1.3 million people prepare to travel to the country for the Muslim pilgrimage, the Hajj. As of 24 August, 62 people were receiving treatment for MERS, and 25 people had died since the start of the month. The coronavirus has claimed 492 lives since arriving in the country in 2012. Saudi officials are considering banning the traditional sacrifice of camels during the Hajj, owing to suspicions that the animals transmit the virus to humans.

Bird blast blunder

Some of New Zealand's rarest birds have been killed in an apparent conservation cull gone wrong. On 21 August,
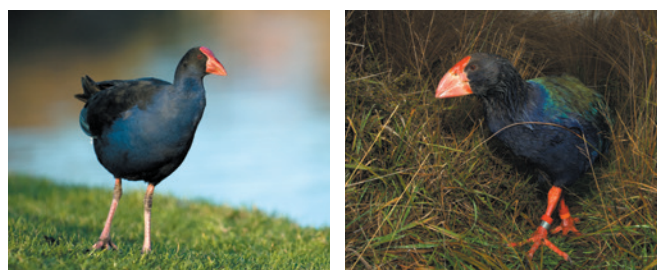
the New Zealand Department of Conservation halted an organized cull of the common pūkeko (*Porphyrio porphyrio*, **pictured, left**) on Motutapu Island after four endangered takahē (*Porphyrio hochstetteri*, **pictured, right**) — which resemble pūkeko — were found dead with shotgun wounds. Twenty-one of the country's 300 takahē lived on the island before the cull, which enlisted the help of local deerstalkers. The president of the New Zealand Deerstalkers' Association apologized to the country, accepting that an investigation into the shooting will place blame on an association member.

Defamation suit

A Massachusetts court has thrown out defamation charges brought by a researcher against the publisher of *Diabetes*. Mário Saad sued the American Diabetes Association in February for publishing 'expressions of concern' about four papers that appeared

in the journal between 1997 and 2011, arguing that the statements had damaged his reputation. But on 18 August the court concluded that the expressions were opinion, and not defamatory. The same court had previously dismissed Saad's request to prevent the journal from printing the statements or retracting the papers. See go.nature.com/v9ymae for more.

FACILITIES

New Scripps centre

The Scripps Institution of Oceanography at the University of California, San Diego, will establish a new centre for climate-change research, thanks to a US$5-million philanthropic donation announced on 24 August. The money was pledged by Carol Dean and Richard Hertzberg, who is president of ENPEX, an energy company involved in both fossil fuels and renewables. The institute,

COMING UP

1–4 SEPTEMBER
The European Society for Translational Medicine gathers in Vienna, Austria, for its third annual congress.
go.nature.com/hone5z

1–5 SEPTEMBER
Cold Spring Harbor Laboratory in New York hosts a meeting of world leaders in eukaryotic DNA replication and genome maintenance.
go.nature.com/lkhp7m

2–9 SEPTEMBER
Scientists in Barcelona, Spain, hop aboard the first Cosmo Cruise, a cosmology conference and Mediterranean cruise combined.
go.nature.com/c2crla

to be called the Center for Climate Change Impacts and Adaptation, will focus on research into coping with the current and future effects of climate change, such as coastal flooding and extreme weather events.

BUSINESS

Goat welfare

A judge at the US Department of Agriculture (USDA) began hearing complaints against one of the world's largest antibody manufacturers on 18 August. California company Santa Cruz Biotechnology has come under fire from the USDA, which has lodged three complaints about the company's treatment of animals — the most that any single company has received. The complaints allege that Santa Cruz Biotechnology hid a facility with 800 goats from the regulators and mistreated sick animals. The hearing was suspended on 21 August.
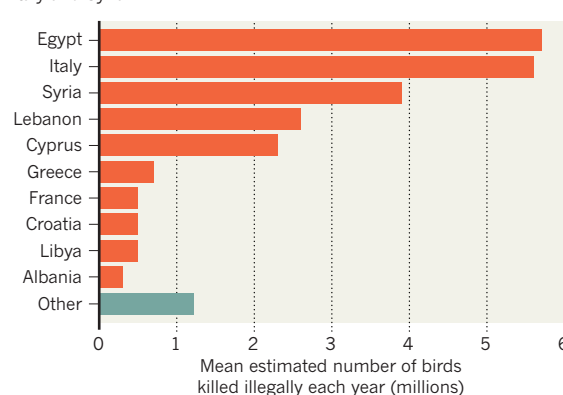
↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Each year in the Mediterranean region, 25 million birds are killed illegally, according to a 21 August report by BirdLife International (see go.nature.com/rqkafx). Egypt, Italy and Syria are the worst countries for illegal bird slaughter. The biggest hotspot is Famagusta in Cyprus, with an annual average of 689,000 killings. The review warns that 40 migratory songbird species — including the European turtle dove (*Streptopelia turtur*) — are declining in numbers, and that illegal killing is at "extraordinary and unsustainable levels".

MEDITERRANEAN BIRD KILL
A survey reveals that 25 million birds are killed illegally in the Mediterranean each year, with the worst offenders being Egypt, Italy and Syria.

Egypt
Italy
Syria
Lebanon
Cyprus
Greece
France
Croatia
Libya
Albania
Other

0 1 2 3 4 5 6
Mean estimated number of birds killed illegally each year (millions)

NEWS IN FOCUS

ERIC GAY/AP

Some of the people who survived Hurricane Katrina lost loved ones, and many were made homeless by the storm.

PSYCHOLOGY

Hurricane Katrina's psychological scars revealed

Mental health worsened in the disaster's aftermath, but survivors also showed resilience.

BY SARA REARDON

New Orleans still bears the scars of Hurricane Katrina, ten years later. More than 500,000 people fled when the storm hit, and many never returned. Large swathes of the city are sparsely populated, particularly in the poor neighbourhoods that suffered the most severe flood damage.

Psychological scars linger, too. Many hurricane survivors continue to experience mental-health problems related to the storm,

whether or not they returned to New Orleans, say researchers tracking Katrina's psychological aftermath. Such work could ultimately aid people affected by future disasters, by identifying factors — such as lack of a social-support network and unstable environments for children — that seem to increase risk of mental-health trauma.

"What's unique about this disaster is the magnitude of it," says Joy Osofsky, a clinical psychologist at Louisiana State University in New Orleans. Katrina, a category 3 hurricane when

it made landfall on 29 August 2005, ultimately damaged an area the size of the United Kingdom. In New Orleans, it destroyed basic resources such as schools and health clinics to a degree unparalleled in recent US history.

Osofsky saw the devastation and despair first hand. With their clinics flooded after the storm, she and other mental-health experts set up treatment centres for emergency responders on cruise ships docked nearby on the Mississippi River, and an emergency psychology unit at the city's central command centre. Osofsky says ▶

that the centres treated thousands of displaced and traumatized people.

Some residents reported watching corpses float by as they waited to escape their flooded homes. Families were separated as emergency responders herded people into temporary holding centres, and police wielding guns prevented newly homeless people from crossing a bridge to escape the city. "There were so many ways in which common humanity was shut down by dystopian images and rumours," says Jean Rhodes, a clinical psychologist at the University of Massachusetts in Boston.

At least one study of storm-ravaged areas in Alabama, Louisiana and Mississippi suggests that mental-health conditions, such as post-traumatic stress disorder (PTSD) and depression, worsened over time[1]. The prevalence of PTSD in these regions rose from 15% a few months after the storm to 21% one year later, and the proportion of people experiencing suicidal thoughts more than doubled from 2.8% to 6.4%. Robert Ursano, a psychiatrist at the Uniformed Services University in Bethesda, Maryland, says that this reflects the difficulties that survivors faced when trying to cope with loss or meet basic needs, such as finding housing.

"There was a level of uncertainty about whether or not recovery would be possible that I don't think has been encountered before in the United States," says Elizabeth Fussell, a demographer who was working at Tulane University in New Orleans when the storm hit.

Rhodes argues that simple steps might have mitigated the mental-health toll. Studies reveal that people who had strong social-support networks when Katrina hit were more resilient to psychological trauma[2]. Yet in many cases, officials' actions severed such ties; many residents of flooded, poor areas were sent by bus to randomly assigned US cities, for example. Sorting evacuees into shelters by neighbourhood could be one way to prevent this stress, says Rhodes.

Rhodes has a rare perspective on the factors that influenced mental and physical health following the storm. In 2003, she and her colleagues had begun to study whether giving college scholarships to a group of 1,019 low-income New Orleans parents would increase their well-being. Katrina halted this work, but the researchers were able to use the already-collected medical and demographic data to track changes in health caused by the hurricane. Such baseline data are rare in disaster research, and are especially valuable for studying poor communities, which tend to have high rates of stress and mental illness.

The renamed, refocused Resilience in Survivors of Katrina (RISK) Project still follows many of the original study participants. Nearly half of the 392 low-income parents participating in the revised project had symptoms of PTSD one year after the hurricane, and the rate of other serious mental illnesses such as depression and psychosis doubled to 14% (ref. 3).

But these storm survivors ultimately displayed resilience. About one-third reported 'post-traumatic growth': the feeling that surviving the disaster made them stronger, even if they simultaneously experienced from mental illness. Three years after the storm, the RISK team found that two-thirds of the 386 African American women who participated in the original study no longer displayed the signs of psychological distress that were evident just after Katrina hit[4].

"People kind of see disasters on TV and say, 'How are these people ever going to build their lives again?'" Rhodes says. "We see that the vast majority of people over time bounce back." ∎

1. Kessler, R. C. *et al. Mol. Psychiatry* **13,** 374–384 (2008).
2. Chan, C. S., Lowe, S. R., Weber, S. R. & Rhodes, J. E. *Soc. Sci. Med.* **138,** 38–43 (2015).
3. Rhodes, J. *et al. Am. J. Orthopsychiatry* **80,** 237–247 (2010).
4. Lowe, S. R. & Rhodes, J. E. *Am. J. Orthopsychiatry* **83,** 398–412 (2013).

North Pacific 'blob' stirs up fisheries management

Unusual warmth strengthens calls for ecosystem–based decisions.

BY VIRGINIA GEWIN

Unprecedented conditions in the Pacific Ocean have sent fisheries managers into uncharted waters. 'The blob', an unusually warm mass of water that has been parked in the northern Pacific for 18 months, has quelled upwelling that delivers nutrients to coastal waters where migratory salmon, tuna and whales fatten themselves on 'forage species' such as anchovies, sardines and krill.

The nutrient shortage comes at a time when populations of many of those forage species are already at historic lows. With a strengthening El Niño — warmth in the eastern equatorial Pacific that affects weather patterns worldwide — fisheries managers face a good deal more uncertainty than usual as they prepare to set catch limits for next year.

The situation lays bare the need to consider ecological processes in fisheries decision-making, scientists said at the American Fisheries Society annual meeting in Portland, Oregon, last week. Managers tend to base limits on assessments that focus on individual species and presume that population trends are stable. Ecosystem-based fisheries management aims for a more comprehensive approach that considers variables such as predator–prey relationships, climate conditions and economic factors.

Up to now, the bodies that set catch limits have used ecosystem-based approaches only sparingly. But last November, the Pacific Fishery Management Council, which makes catch recommendations for the West Coast to the US National Marine Fisheries Service, reviewed and conditionally endorsed a comprehensive ecosystem-based computer model — the first step towards bringing such a tool into management decision-making.

"Uptake of these principles has been relatively slow," says Tim Essington, a fisheries scientist at the University of Washington in Seattle.

The biggest barrier may be the need to collect and analyse relevant biological data, such as information on how fluctuations in the population of a prey species will affect its predators. Supporters of ecosystem-based fisheries are creating tools such as the California Current Predator Diet Database, which is amassing information about the eating habits of 119 species. At the fisheries meeting, Amber Szoboszlai, a research analyst at the Farallon Institute for Advanced Ecosystem Research in Petaluma, California, showed how she had used the database to determine that fish eat 75% of the anchovies consumed in the Pacific, whereas mammals eat only 16% and seabirds 7%.

If the blob signals a regime change in the Pacific, conservationists argue that ecosystem-based management will be essential to preventing a catastrophe. "The whole system seems to be changing radically," says Rebecca Goldburg, director of ocean science for the Pew Charitable Trusts in Washington DC. ∎

CLIMATE

Minnesota bog study turns up the heat on peat

Experiment boosts temperature and carbon dioxide to gauge global-warming response.

BY ALEXANDRA WITZE

In a peat bog in northern Minnesota, researchers have built a little patch of the future.

Ten octagonal enclosures — each 12 metres across, 8 metres high and open to the sky — dot the US Forest Service's Marcell Experimental Forest. Boardwalks wend their way around them, allowing scientists to walk above a pristine 8.1-hectare peat bog and get a close-up peek at Earth's greenhouse fate.

Half of the enclosures have carbon dioxide gas wafting into them at concentrations of 900 parts per million — more than twice that of today's atmosphere. Hot-air blowers and soil heaters raise temperatures by up to 9 °C above that of the outside air. The combination of warmth and elevated CO_2 places this study in a wave of experiments that are probing how some of Earth's most sensitive ecosystems — in this case, carbon-rich peatlands — will respond to climate change.

The US Department of Energy and the Forest Service were this week set to formally kick off the full US$10.5-million experiment, called Spruce and Peatland Responses Under Climatic and Environmental Change (SPRUCE). The study is planned to run for a decade.

"We think this gives us a good glimpse as to what future climates might look like," says Paul Hanson, an ecologist at Oak Ridge National Laboratory in Tennessee and the experiment's project coordinator. "It allows us to cheat today's conditions and take on the whole ecosystem, from the top of the trees to the deep peat."

Previous experiments have typically raised either CO_2 levels or temperatures in small sample plots of ecosystems, such as temperate and tropical forests. SPRUCE is rare in using both elevated CO_2 and heating on a large scale, and in warming the soil to 2 metres' depth.

Although they cover just 3% of Earth's surface, peatlands contain roughly one-third of the planet's terrestrial carbon stores. As temperatures rise, bogs are expected to release more carbon, contributing to even

"It allows us to cheat today's conditions and take on the whole ecosystem."

Study enclosures will be monitored continuously over a decade.

greater warming. But computer models of future climate do not accurately account for peatland carbon emissions, Hanson says. In the SPRUCE experiment, scientists do not yet know whether that carbon release will come mostly in the form of CO_2 or as the more potent heat-trapping gas methane. Although bogs release mainly CO_2 today, methane could become the primary greenhouse gas produced by the SPRUCE bog as it warms — as long as the peat stays wet. If it dries out, the types of microbe living in the peat may shift and begin putting out more CO_2 again.

"Experiments like SPRUCE not only let us test what we think we know — often they surprise us by showing us what we don't," says Susan Trumbore, a biogeochemist at the Max Planck Institute for Biogeochemistry in Jena, Germany.

HOT PLATES

In June 2014, SPRUCE technicians flipped a switch to turn on resistance heaters that penetrate the peat bog in concentric circles. As a result, 2 metres down, carbon-rich decaying moss that dates back 4,000–7,000 years is now being heated to as much as 9 °C above the temperature of the outside air. The experiment's air heaters were switched on in mid-August.

Researchers have already seen hints of how the peat responds to warming. Water flowing through the bog, for instance, has not changed much chemically in the past year, says Natalie Griffiths, an aquatic biogeochemist from Oak Ridge who presented the findings this month at a meeting of the Ecological Society of America. That suggests the peat may be fairly resistant to warming, she says — at least at the start. Long-term studies will include tests of whether soil microbes — and their roles in the ecosystem — change as temperatures rise.

What SPRUCE finds in this northern boreal forest should inform permafrost-warming studies that the US Department of Energy and others have been carrying out across the Arctic. More broadly, SPRUCE could unite with other soil experiments in an international network that would exchange results. Hanson has joined Margaret Torn of California's Lawrence Berkeley National Laboratory and others in calling for an international link-up to share data on how soils might respond to changing climate (M. S. Torn *et al. SOIL* **1,** 575–582; 2015).

Despite its goals, SPRUCE itself has a sizeable carbon footprint. The electricity bill soars to more than $1,000 a month in the summer, and the experiment will be even more costly to run through the cold Minnesota winter. If ambient temperatures fall low enough — to about −26 °C — the propane gas running through the heating pipes will no longer be able to maintain the necessary warmth. "We're planning to run year-round," says Hanson, "unless it's so cold the engineering portion of the system can't keep up." ■

BIOTECHNOLOGY

Biohackers gear up for genome editing

Amateurs are ready and able to try the CRISPR technique for rewriting genes.

BY HEIDI LEDFORD

A complete lack of formal scientific training has not kept Johan Sosa from dabbling with one of the most powerful molecular-biology tools to come along in decades.

Sosa has already used CRISPR, a three-year-old technology that makes targeted modifications to DNA, in test-tube experiments. Next week, he hopes to try the method in yeast and, later, in the model plant *Arabidopsis thaliana*.

Hailed for its simplicity and versatility, CRISPR allows scientists to make specific changes to a gene's sequence more easily than ever before. Researchers have used CRISPR to edit genes in everything from bacteria to human embryos; the technique holds the potential to erase genetic defects from family pedigrees plagued by inherited disease, treat cancer in unprecedented ways or grow human organs in pigs. One researcher has even proposed modifying the elephant genome to produce a cold-adapted replica of the long-extinct woolly mammoth.

Such feats are beyond the reach of do-it-yourself (DIY) 'biohackers', a growing community of amateur biologists who often work in community laboratories, which typically charge a recurring fee for access to equipment and supplies. But CRISPR itself is not. Driven by an inventive spirit that inspires them to fiddle with yeast to alter the flavour of beer, build art installations out of bacteria or pursue serious basic-research questions, these amateurs cannot wait to try the technique.

"It's, like, the most amazing tool ever," says Andreas Stürmer, a biohacker and entrepreneur who lives in Dublin. "You could do it in your own home."

Sosa is an IT consultant from San Jose,

Biohackers such as Johan Sosa are exploring the creative potential of molecular biology.

California, who took up biohacking as a hobby about three years ago, when he decided that he would like to grow organs — or maybe other body parts — in the lab. At first, he had no idea how unrealistic that goal was. "I just thought you take a bunch of stem cells and add stuff to them," he says.

The challenge of manipulating living cells sank in as he began to read molecular-biology textbooks, attend seminars and teach himself laboratory techniques. He joined the BioCurious community lab in Sunnyvale, California.

Sosa is not quite sure what he will do with CRISPR once he has mastered it. He might participate in a group effort at BioCurious to engineer yeast to produce casein, a protein found in milk, as a step towards making vegan cheese. That could involve using CRISPR to learn how proteins are chemically modified in different types of yeast. "Now we have this ability to do what the major labs have been doing all this time," he says. "It's very exciting."

Artist Georg Tremmel, a research fellow in biological-data visualization at the University of Tokyo, has clear plans for CRISPR. He and his collaborators plan to 'de-engineer' genetically modified blue carnations sold in Japan by snipping out the inserted gene that turns the flower blue, thus reverting it to its 'natural', white state. They want audiences to ponder whether these doubly modified carnations should be deemed any different from unengineered plants with essentially the same genome.

So far, the hardest part of the project has not been using CRISPR, but growing the
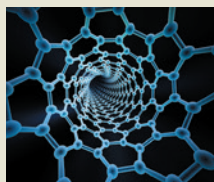
---

carnations in cell culture, Tremmel says. Another challenge will be getting permission to exhibit the work: although the blue carnations have been approved for sale in Japan, the de-engineered white carnations may need regulatory approval before they can be taken out of a laboratory.

In addition to its creative possibilities, CRISPR also poses potential for mischief. The US Federal Bureau of Investigation's Bioterrorism Protection Team has painstakingly forged relationships with the biohacker community over the past few years and regularly reminds its members to keep an eye out for suspicious activity. Those concerns may be unnecessary, says Todd Kuiken, who studies science policy at the Wilson Center, a think tank in Washington DC. Most biohackers have benign goals, he says, such as creating rainbow-coloured bacteria or brewing distinctive beer.

There is also a tendency to overestimate what a typical DIY biologist can do, Kuiken adds. Reagents such as enzymes and antibodies are expensive, molecular-biology experiments are time-consuming and equipment that professional scientists take for granted is often beyond the means of individuals or community labs. And most community labs insist that their members work only with organisms that require the lowest level of biosafety precautions, which leaves human cells and most pathogens off the menu. In some parts of Europe, genetic engineering is illegal outside of professional facilities.

⟳ **NATURE.COM**
CRISPR: The good, the bad and the unknown
nature.com/crispr

Given the constraints of a DIY lab, many hobbyists resort to CRISPR only when they need an extremely precise change to the genome, says Keoni Gandall, a 16-year-old biohacker and science-fair champion from Huntington Beach, California, who has been working with polymerase chain reaction machines and centrifuges at home for about three years. So far, Gandall has used CRISPR only while volunteering in a local university lab. "It's pretty good," he says.

One of the biggest fears surrounding CRISPR is that it could be used to create a genetic modification designed to spread through a population of organisms at an unnaturally fast rate. But Dan Wright, an environmental lawyer and DIY biohacker in Los Angeles, California, thinks that such a scenario is still beyond the ability of most amateurs. Constructing such a system would surpass the relatively simple tweaks that he and his colleagues are contemplating.

"It's too difficult," Wright says. "Just knocking out a gene in one plant is enough of a challenge for a biohacker space at this point." ∎

A tree grows in Baltimore: the city is at the heart of the burgeoning field of urban ecology.

ECOSYSTEMS

Ecologists embrace their urban side

Climate change and the rise of cities have broadened what it means to study ecosystems.

BY DANIEL CRESSEY, BALTIMORE, MARYLAND

A concrete megalith overshadowed by skyscrapers and surrounded by roads that roar with traffic, the convention centre in downtown Baltimore may seem an inappropriate setting for an ecology conference. But the resolutely urban backdrop for the annual meeting of the Ecological Society of America (ESA) is a fitting symbol of the growing acceptance of, and interest in, 'urban ecology' — the study of cities and the organisms that dwell in them as ecosystems.

"In the past 10 years, it's really become more mainstream. People's reactions have shifted from 'What's that?' or 'Why do you do that?' to 'Oh, cool,'" says Laura Martin, a historian and urban ecologist at the Harvard University Center for the Environment in Cambridge, Massachusetts. She presented work at this year's centennial meeting (which ran on 9–14 August) showing that orange jewelweeds (*Impatiens capensis*) in Manhattan and other urban settings are evolving defences to incursions of certain deer that eat them (L. J. Martin *et al. J. Ecol.* **103,** 243–249; 2015). Martin is part of a team that called at ▶

▶ the 2010 ESA meeting for more researchers to resist ecology's traditional bias towards pristine ecosystems (*Nature* http://doi.org/fwd62r; 2010).

This now seems to be happening. The sheer number of presented papers relating to cities this year points to a "phenomenal" growth of urban ecology, says Wayne Zipperer of the US Department of Agriculture Forest Service, who studies the urbanization of ecosystems in Gainesville, Florida. He recalls an ESA conference in the early 1990s that had just one oral presentation and one poster on the sub-field — this year, there were around 450 presentations, posters and events that touched on urban issues, roughly 10% of the conference total. Among them: a study of people in New York City, in particular immigrants from China, who rely on fungi and street trees such as ginkgo, white oak and linden for foods and medicines; and research showing that urban streams are as microbially diverse as wild ones.

The rise in papers shows "the maturity of urban ecology", says Mark McDonnell, director of the Australian Research Centre for Urban Ecology in Melbourne and editor of the new *Journal of Urban Ecology*, which started accepting submissions in March.

↻ **NATURE.COM**
The science of cities: Life in the concrete jungle
go.nature.com/ehu1wb

Urban ecologists attribute the swell of interest in their discipline to multiple factors, including the realization that human actions are warming the planet, that people are migrating to cities in increasing numbers and evidence that the study of urban ecosystems provides important and practical insights.

One move in particular that spurred the field was the decision in 1997 by the US National Science Foundation (NSF) to include two cities — Baltimore and Phoenix, Arizona —

"Standing room only for urban ecology? When does that happen?"

in a group of more than 20 long-term ecological research (LTER) sites that it funds. The studies based on these sites "changed the way ecologists feel about working in urban situations", says Steward Pickett, an urban ecologist at the Cary Institute of Ecosystem Studies in Millbrook, New York, and director of the Baltimore LTER site.

The studies are offering up results. Among many findings, the Baltimore Ecosystem Study showed that urban streams — historically considered useless dead zones — retain nitrogen run-off from fertilizers, providing a valuable 'ecosystem service' by preventing the nitrogen from reaching other water courses where it can spawn damaging algal blooms.

Such findings are becoming ever more important as both the absolute number of people and the proportion of the world's population living in cities grows, and as cities seek resilience to the effects of climate change, said sustainability scientist Nancy Grimm to meeting attendees. In July, researchers including Grimm — who heads the Phoenix LTER project from Arizona State University in Tempe — launched the Urban Resilience to Extremes Sustainability Research Network, a US\$12-million project funded by the NSF. The aim is to work out how cities can be designed to better withstand storms such as Hurricane Sandy, which hit New York in 2012, and Hurricane Katrina, which tore apart New Orleans in 2005. Such extreme weather events are predicted to become more frequent in future.

The behaviour of ecologists at the meeting suggests that they may already be converts. As one session started, Madhusudan Katti, an ecologist at California State University in Fresno, looked around as people continued to file into the overcrowded room. "Standing room only for urban ecology?" he said. "When does that happen?" Jon Christensen, a historian of science and the environment at the University of California, Los Angeles, detects another sign that a shift has occurred. "The city is no longer 'the other', the negative example contrasted with the pristine," he says. "Urban ecology has arrived." ∎
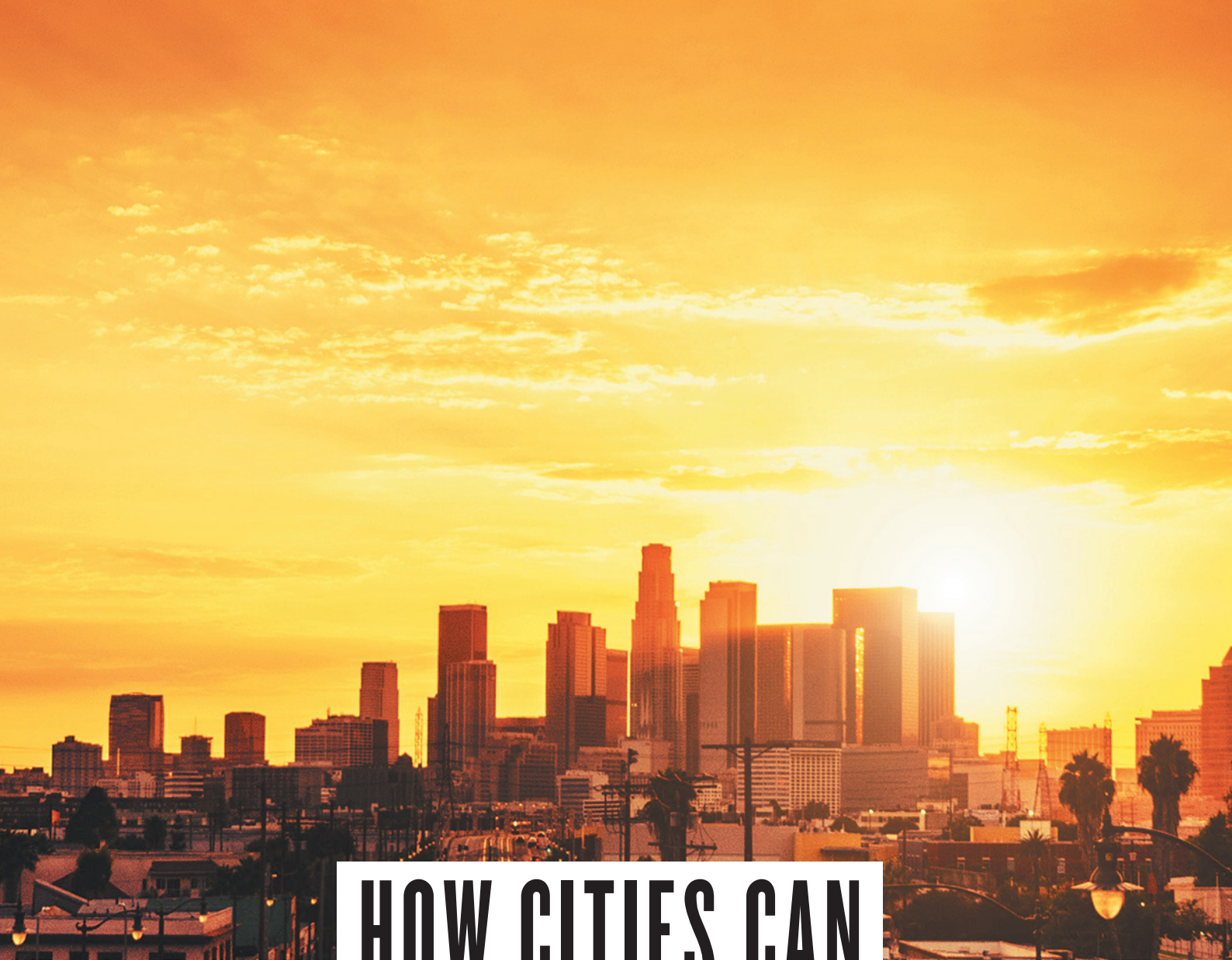
HOW CITIES CAN
BEAT THE HEAT

Rising temperatures are threatening urban areas, but efforts to cool them may not work as planned.

BY HANNAH HOAG

The greenhouses that sprawl across the coastline of southeastern Spain are so bright that they gleam in satellite photos. Since the 1970s, farmers have been expanding this patchwork of buildings in Almería province to grow produce such as tomatoes, peppers and watermelons for export. To keep the plants from overheating in the summer, they paint the roofs with white lime to reflect the sunlight.

That does more than just cool the crops. Over the past 30 years, the surrounding region has warmed by 1 °C, but the average air temperature in the greenhouse area has dropped by 0.7 °C (ref. 1).

It's an effect that cities around the world would like to mimic. As Earth's climate changes over the coming decades, global warming will hit metropolitan areas especially hard because their buildings and pavements readily absorb sunlight and raise local temperatures, a phenomenon known as the urban heat island effect. Cities, as a result, stand a greater chance of extreme hot spells that can kill. "Heat-related deaths in the United States outpace — over the last

Last year, Los Angeles ruled that new and renovated homes must have 'cool roofs'.

30 years — all other types of mortality from extreme weather causes," says Kim Knowlton, a health scientist at Columbia University in New York. "This is not an issue that is going away."

Some cities hope to stave off that sizzling future. Many are planting trees and building parks, but they have focused the most attention on rooftops — vast areas of unused space that absorb heat from the Sun. In 2009, Toronto, Canada, became the first city in North America to adopt a green-roof policy. It requires new buildings above a certain size to be topped with plants in the hope that they will retain storm water and keep temperatures down. Los Angeles, California, mandated in 2014 that new and renovated homes install 'cool roofs' made of light-coloured materials that reflect sunlight. A French law approved in March calls for the rooftops of new buildings in commercial zones to be partially covered in plants or solar panels.

But the rush to act is speeding ahead of the science. Although cool roofs and green roofs can strongly curb temperatures at the tops of buildings, they do not always yield benefits at the street level, and they may trigger unwanted effects, such as reducing rainfall in some places. "There was a notion that the community had reached a conclusion and there was a one-size-fits-all solution," says Matei Georgescu, a sustainability scientist at Arizona State University in Tempe. "But that is not the case."

On top of that, it is unclear whether the limited programmes currently in place will have a measurable effect on temperature — and citizen health — and whether cities will expand their efforts enough to produce results. "If you're just putting green roofs on city hall and schools, it's not going to move the needle," says Brian Stone Jr, an urban scientist at the Georgia Institute of Technology in Atlanta.

HOT TIME IN THE CITY

For ten days in August 2003, an unprecedented heatwave stifled Western Europe, breaking records reaching back five centuries (see *Nature* http://doi.org/fvgkt4; 2004). Daytime temperatures in Paris shot up to 40 °C and nights remained torrid. By the end of August, the death toll from dehydration, hyperthermia, heat stroke and respiratory problems for all of Europe surpassed 70,000, with many fatalities in the urbanized areas around Paris and Moscow.

This is just a taste of conditions to come. Regional climate models indicate that by 2050, week-long heat spells on par with the August 2003 event may strike once a decade in Eastern Europe and every 15 years in Western Europe[2]. Across the globe, the number, duration and frequency of heatwaves is projected to increase. "This is one of the few extreme events where all of the models agree with each other," says Dan Li, a climate modeller at Princeton University in New Jersey.

And when temperatures rise, cities suffer disproportionately because of the way they are built. Dark roofs, roads and other construction materials absorb incoming short-wave radiation from the Sun and re-radiate it as long-wave energy, warming the atmosphere nearby. Air conditioning adds to the problem by pulling heat from inside buildings and vehicles and dumping it outside, further driving up urban temperatures.

In the absence of interventions, heat islands will only grow: by 2050, urban surface area in the United States is expected to expand by one-third. At the same time, the global population is projected to grow to 9.6 billion, with two-thirds living in urban areas, compared with just over half today. It all adds up to more heat-trapping potential and more people affected by extreme heat. And yet despite the risks, few cities have plans in place to address urban heat directly. In the United States, says Stone, "most cities are ignoring the climate issue".

Los Angeles is not. The city has seen its annual average temperature rise by more than 2 °C since 1878. By mid-century, its downtown area could face 22 days of extreme heat annually (temperatures exceeding 35 °C), nearly four times the long-term average[3]. To counter the warming, the city aims to convert 10,000 dark-coloured roofs to cool roofs by 2017. By pairing this effort with street plantings and reflective pavements, it intends to shave 1.65 °C off the urban heat island effect by 2035.

Chicago, Illinois, has also become a leader on this issue: it hopes to prevent the kind of mass deaths seen during the city's 5-day heatwave in 1995,

when 700 people died. Since that disaster, it has added cool roofs, green roofs and street plantings — and transformed black-top playgrounds into grass fields. Incentives have helped to trigger the construction of more than 516,000 square metres of green roofs on 509 buildings.

Toronto is rapidly catching up. It requires new buildings taller than six storeys and with more than 2,000 m^2 of roof space to cover 20–60% of that with plants. Since 2010, the city has added 260 green roofs covering 196,000 m^2.

Some forms of cool roofs can be comparable in price to regular ones, but green roofs are more expensive to install, and they have higher maintenance costs. They offer other benefits, however, such as slowing storm-water run-off, providing habitat for pollinating insects and making cities more beautiful.

SEEING GREEN

On the roof of the University of Toronto's architecture building, bumblebees flit from one yellow flower to another. Located in the city's downtown, the building is topped by a patchwork of 33 rectangular, raised garden beds planted with native grasses, flowers or non-native sedums — plants with waxy, water-storing leaves. Each bed has a different combination of plants, soil and irrigation techniques — all of which are monitored by 270 sensors measuring air temperature, soil temperature, soil moisture and rainwater run-off. The garden beds are part of the Green Roof Innovation Technology Laboratory (GRIT Lab), the only facility of its kind in Canada to test the performance of green roofs and other strategies to mitigate climate change.

Green roofs reflect more sunlight than conventional tar or gravel roofs, but they get much of their cooling power from moisture in the plants and soil. As water in leaves and soil evaporates, it carries heat to the atmosphere and lowers air temperature nearby, just as athletes cool off when their sweat evaporates. Compared with a black roof, a green roof can be 40 °C cooler on a hot summer day. Green roofs also act as insulators and reduce energy costs associated with cooling.

The coolest test bed on the GRIT Lab roof is irrigated and contains organic soil and a thick mat of sedum that flows over the box edges. Its neighbour has a patchy lawn of meadow grasses growing in an unirrigated box lined with a porous rocky medium that is widely used on green roofs. At its surface, the sedum box is 4 °C cooler than the air temperature, but the sparsely planted meadow-grass box can be 14 °C hotter, notes Liat Margolis, GRIT Lab's director. "You might as well not have it on the roof," she says. Experiments such as this show how difficult it can be to find the right combination of substrate, vegetation and irrigation to have an impact on the temperature of one rooftop, Margolis adds.

And even with the best green roofs, no one knows how much this approach can cool a whole city. Only a few simulations have evaluated green roofs at that scale. A decade-old report prepared for the city of Toronto suggests that if the city's entire 50 square kilometres of available rooftop were converted to green roofs, the ambient air temperature would reduce by 0.5–2 °C. But the area of green roofs added since the by-law came into effect amounts to less than 0.5% of the city's available roof space.

In the Baltimore–Washington metropolitan area, a 2014 study projected that 90% of the rooftops had to be planted to decrease the daily maximum air temperature by 0.5 °C during a 3-day heatwave[4].

> "IF YOU'RE JUST PUTTING GREEN ROOFS ON CITY HALL AND SCHOOLS, IT'S NOT GOING TO MOVE THE NEEDLE."

Because there is less evaporation when the Sun goes down, green roofs do not cool as well at night — a deadly time during heatwaves. And rooftop plantings release stored heat after the Sun goes down. "If you try to mitigate the urban heat by putting up green roofs, it will do some good for reducing temperatures during the day, but it might increase at night," says David Sailor, an urban climate scientist at Portland State University in Oregon.

It's also unclear how well green roofs many storeys up offer relief to people below. A modelling study[5] that replicated a day during the 2003 European heatwave found that planting 25% of the roof area along a street in the centre of Arnhem in the Netherlands had no effect on street-level temperatures because the wind blew away the cooler air before it could reach the ground.

Observations of green roofs in Chicago have also raised questions about their benefits. Comparing satellite images from 1995 and 2009, researchers checked how surface temperatures had changed at spots within the city where dark roofs and pavement had been replaced by vegetation or brighter coatings.

The green roofs did not alter temperatures significantly, but cool roofs did[6]. They increased the city's reflectivity, or albedo, by 1.6% — equivalent to the cooling power of 65,000 large, window-sized air conditioning units operating at full capacity over the summer. Planting trees and converting paved areas to grass were also more effective than green roofs.

DOWNSIDES OF COOL

The effects in Almería are even bigger. Each summer, after farmers whitewash the greenhouse roofs with slaked lime, they reflect 35% of the incident sunlight. The pastures nearby reflect just 15%, on par with most cities.

Simulations show that by whitening roads and roofs, cities could cool down considerably. Doubling the rooftop albedo of Los Angeles, which is on average 17%, could trim temperatures by 0.5 °C in some areas and by as much as 2 °C in others, says Haider Taha, an urban atmospheric modeller and president of Altostratus, a meteorological company in Martinez, California. "One degree Celsius isn't too much to ask for, and according to the models, it's doable," he says.

But cool roofs could also produce some unfavourable effects, depending on the location. If the albedo is increased too much, it could slow local sea breezes, reduce air quality or warm downwind areas, says Taha. "Each city has a threshold, and if you go beyond that, some things start looking bad."

And cool roofs could also inhibit rain. In many regions, heating of the ground during the day causes moist air to rise, driving cloud formation and precipitation. "If we don't have that, then we don't have precipitation," says Georgescu. In a modelling study[7], he found that if cool roofs were widely implemented in urban areas from Florida to the north-eastern United States, daily summertime precipitation could decrease

Researchers test different types of vegetation in Toronto, Canada.

by 2–4 millimetres by 2100.

Despite the uncertainties, many scientists say that cities are not pursuing cooling strategies quickly enough, given the pace of climate change and urban growth. Proponents say that both green roofs and cool roofs have helped in some situations and that careful implementation could improve their efficiency. Stuart Gaffin, an urban-climate scientist at Columbia University in New York, warns against placing too much stock in modelling studies that forecast unwanted side effects such as reduced cloud cover and rainfall. Clouds are among the most complex things to model, and cities already enhance rainfall because of the particulates they produce, he says.

Despite all of the heat-related risks that cities face in the future, few have put heat-management plans in place. Louisville in Kentucky is one: it will soon become the first major US city to develop an urban heat-adaptation plan, says Stone, who is leading the project. The effort is driven by necessity. Louisville has the fastest warming urban heat island in the United States, and temperatures there have climbed by more than 4 °C since 1961. Part of the problem is that the city has lost 54,000 trees per year to insects, ice storms and lack of care.

Stone is now collecting the baseline data that most cities lacked before embracing cooling steps. He is travelling around Louisville measuring tree cover, finding hot spots and identifying areas with vulnerable residents. The next step is to create a blueprint that combines cool roofs, green roofs, tree plantings and cool paving materials that could change the fate of the city's most at-risk residents. Stone is starting with modest but realistic assumptions in his modelling: the conversion of just 100 buildings to green roofs, for example. At the same time, the city hopes to increase its number of trees.

If Louisville implements the strategies that Stone recommends, it could become a testing ground that will reveal how changes to a city's physical surface alter the urban heat island — and its pioneering programme could point the way for other cities to follow. "We're already crossing thresholds that are pretty sensitive," says Stone. "Cities are going to be contemplating more aggressive action. But cities can measurably slow the rate at which they're warming over a decade or two." And that's pretty quick, he adds, because even if we eliminated greenhouse-gas emissions tomorrow, "we're still going to warm for a couple of hundred years". ∎

Hannah Hoag *is a freelance journalist in Toronto, Canada.*

1. Campra, P. & Millstein, D. *Environ. Sci. Technol.* **47,** 12284–12290 (2013).
2. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García–Herrera, R. *Science* **332,** 220–224 (2011).
3. Sun, F., Walton, D. B. & Hall, A. *J. Climate* **28,** 4618–4636 (2015).
4. Li, D., Bou-Zeid, E. & Oppenheimer, M. *Environ. Res. Lett.* **9,** 055002 (2014).
5. Gromke, C. *et al. Build. Environ.* **83,** 11–26 (2015).
6. Mackey, C. W., Lee, X. & Smith, R. B. *Build. Environ.* **49,** 348–358 (2012).
7. Georgescu, M., Morefield, P. E., Bierwagen, B. G. & Weaver, C. P. *Proc. Natl Acad. Sci. USA* **111,** 2909–2914 (2014).

GRIT LAB

TICK TROUBLE

*Scientists have no shortage of ideas
about how to stop tick–borne illnesses.
What is holding them back?*

BY MELINDA WENNER MOYER

On a balmy day in late June, Scott Williams waits for a white-footed mouse (*Peromyscus leucopus*) to fall asleep. Williams, a wildlife biologist with the Connecticut Agricultural Experiment Station in New Haven, has just transferred the animal from a trap to a plastic bag containing a cotton ball doused in anaesthetic. As soon as the mouse's breathing slows to one breath per second, Williams will take it out, draw blood, weigh it, put an ear tag on it for identification and check the animal for ticks, saving any that are engorged with blood. He must work quickly. The mouse will wake up in about two minutes, and she might be grumpy.

Williams is testing whether vaccinating mice against *Borrelia burgdorferi*, the bacterium that causes Lyme disease in the United States, can reduce the proportion of ticks that are infected. Health officials are looking on with interest. Connecticut has one of the highest rates of human Lyme disease in the country, and June is peak time for transmission. *Borrelia burgdorferi* infects an estimated 329,000 people in the United States each year, according to the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. And although most people who get prompt treatment recover quickly — Williams has had Lyme three times — up to one in five develops long-term and potentially life-threatening symptoms, including heart, vision or memory problems, or debilitating joint pain.

Williams's approach is one of several strategies being tested in an attempt to thwart the spread of tick-borne diseases. Some, like the mouse vaccine, interrupt the pathogen's ecological circuitry by targeting the wild animals that pass along and amplify the disease. Others, such as efforts to revive a human Lyme vaccine,

aim to protect people from infection directly. A more radical approach could hamper the ability of ticks to bite humans or animals, potentially protecting against dozens of illnesses spreading across the United States, Europe, Africa and Asia.

That the field needs creative solutions is clear. Many long-recommended interventions, such as pesticide application or controlling populations of deer, which are an important host for adult ticks, have had mixed success in scientific studies. Even the time-honoured protective strategies that most people use are not evidence-based. "We tell people to wear repellents, to do tick checks and to shower if they've been in the field, but there's very little data to show that these things reduce human illness," explains Ben Beard, chief of the CDC's bacterial-diseases branch in the division of vector-borne diseases.

Diseases spread by ticks are on the rise around the world, spurred by a combination of factors, including shifting climates and population sprawl into rural areas. Reported cases of Lyme, the most common US tick-borne illness, have nearly tripled in the country since 1992, although some of the increase could be due to heightened awareness. Lyme is also a growing problem in parts of Europe, Mongolia and China. Yet as bad as it is, there are nastier threats on the rise. In parts of Africa, the Middle East, Asia and southern Europe, ticks can spread Crimean–Congo haemorrhagic fever, which is fatal in 40% of cases. And a tick-borne relapsing fever afflicts as many as 1 in 20 residents in parts of Senegal. In the United States, ticks spread at least 16 illnesses, including anaplasmosis, babesiosis, ehrlichiosis and Rocky Mountain spotted fever, all "serious, life-threatening infections", Beard says. And many are increasing in

incidence more quickly than Lyme. In a July 2015 position statement, the Entomological Society of America argued for a national strategy to combat tick-borne diseases. "The recent confluence of environmental, ecological, sociological, and human demographic factors," it said, "has created a near 'perfect storm' leading to more ticks in more places throughout North America."

BACKYARD BATTLEGROUNDS

Williams tags, weighs and releases his mouse just in time. It has no ticks to bring back to the lab for further analysis, but there will be other opportunities. Members of 32 Connecticut households have volunteered to place traps around their properties, and some will also get boxes of mouse treats laden with vaccine. The hope is that, over time, fewer mice and ticks will harbour the bacteria at the sites with the vaccine bait.

The plan is unconventional, because most Lyme-control measures focus on white-tailed deer (*Odocoileus virginianus*), which have exploded in number in the United States over the past century as young forests have become increasingly fragmented by human development and large predators have been all but eradicated. Adult blacklegged ticks (*Ixodes scapularis*) typically feed and mate on deer, so many scientists have argued that the only way to get rid of Lyme is to get rid of the deer.

But such efforts have had "an incredibly spotty record", says Richard Ostfeld, a disease ecologist at the Cary Institute of Ecosystem Studies in Millbrook, New York, who has been studying tick-borne diseases for decades.

When Sam Telford, an epidemiologist at Tufts University in North Grafton, Massachusetts, and his colleagues cut the deer population on

The body of an adult ixodid tick expands as it feeds.

Great Island in Cape Cod by 50% in the early 1980s, they saw no drop in tick numbers — the number of tick larvae on the island actually increased[1]. Ostfeld argues that you do not need many deer to maintain a large tick population. When deer numbers drop, ticks can either crowd in on the remaining deer or find other hosts. Only when almost all of the deer on Great Island had been eliminated did tick populations plummet. But, says Telford, "it is a nightmare trying to get the deer population down that low". And anywhere that is not an island, keeping populations down is practically impossible.

DANGER MOUSE

Ostfeld and others contend that mice are a major driver for both the tick problem and the disease problem. Mice, like deer, flourish in fragmented woodlands — in part because predators such as foxes and opossums get displaced. Ticks then thrive on the rodents, which are poor groomers. Studies suggest that larval ticks have a 50% chance of surviving when they feed on mice, but only a 3.5% chance on opossums[2].

And mice are typically where ticks pick up *B. burgdorferi*. Most mice in Lyme-endemic areas get infected with the bacterium at a young age and, for reasons that are not completely clear, they are particularly good at transmitting it to ticks. Almost all young ticks that feed on white-footed mice become infected, compared with a mere 1% of ticks that feed on deer. Interrupting the tick–mouse infection cycle, says Ostfeld, could make ticks a lot less dangerous.

Maria Gomes-Solecki, a medical microbiologist at the University of Tennessee Health Science Center in Memphis, agrees — which is why she invented the mouse vaccine that Williams is testing. It primes the mice to make antibodies against outer surface protein A (OspA), a molecule that *B. burgdorferi* expresses when it is in a tick's gut. A mouse eats the vaccine, then starts to produce OspA antibodies. The next time a tick feeds on the mouse, the antibodies attack the bacteria in its gut, clearing the infection. As the proportion of ticks infected with *B. burgdorferi* drops, it becomes less likely that the next generation of mice will pick up the parasite, even without vaccination.

Ostfeld and his colleagues reported the first field tests[3] of Gomes-Solecki's vaccine in 2014, and found that although only 28% of the mice in an area that they targeted for 5 years developed protective levels of OspA antibodies, the prevalence of infected blacklegged-tick nymphs (the life stage between larvae and adults) dropped by 75%. The bait-based vaccine is also attractive because it is less ecologically destructive than other strategies — it does not kill animals or even ticks, just the pathogens.

Gomes-Solecki, who licensed her technology to a company she founded, US Biologic in Memphis, would like to see homeowners putting walk-through bait boxes for mice around their gardens. Or, she says, local governments could disperse the bait in parks or forests, much as they do with bait-based rabies vaccines for raccoons and coyotes. "The rodents seem to love them," Williams says of the vaccine-laced treats. One of his colleagues calls them "Fritos for mice".

"WE'VE DISRUPTED THE BALANCE OF NATURE."

Other scientists argue for more a direct means of protecting people against Lyme, ideally with a human vaccine. When vaccine researcher Stanley Plotkin's son was 35, he fell ill with Lyme disease. As often happens with the infection, a doctor missed the diagnosis and the young man went untreated for months. Bacteria invaded his heart and he collapsed one day while walking his dog. Plotkin, now an emeritus professor at the University of Pennsylvania in Philadelphia, says that when paramedics arrived, his son's heart rate was dangerously low. He has since recovered, but the experience "further convinced me, if I needed any convincing", Plotkin says, "that the lack of a Lyme-disease vaccine was a public-health tragedy".

Plotkin worked on a vaccine in the 1990s. Ultimately, a competing product called LYMErix, manufactured by UK-based pharmaceutical company SmithKline Beecham (now GlaxoSmithKline), was approved by the US Food and Drug Administration in 1998. It reduced the risk of Lyme caused by US strains of *Borrelia* by 76% in clinical trials[4]. But it faced problems from the start. First, it garnered lukewarm support from health officials in the United States and was recommended only for people aged 15 to 70 in regions where Lyme is endemic. Then, some recipients complained of autoimmune-related side effects such as arthritis and filed lawsuits against SmithKline Beecham. The company voluntarily shelved LYMErix in 2002. Plotkin maintains that this was a mistake. "The vaccine was safe," he says.

Now, a new and potentially improved vaccine has completed safety trials[5]. Developed by researchers at Stony Brook University and Brookhaven National Laboratory in New York, and licensed to Baxter Innovations in Vienna, the vaccine is similar to LYMErix in that it targets OspA, but it does not contain the protein segment that some scientists and consumers feared could cause an autoimmune reaction. It also contains several variants of OspA, so it protects against many *Borrelia* species known to cause Lyme in humans, including those that affect people in Europe.

Nevertheless, the vaccine's future is uncertain: in 2014, Pfizer bought the rights to sell many of Baxter's vaccine products, but not the Lyme candidate. Baxter is now in talks with Great Plains Biotechnology of Roca, Nebraska, which has expressed interest in purchasing and developing the Lyme vaccine.

Richard Marconi, a microbiologist and vaccinologist at Virginia Commonwealth

Disease ecologist Rick Ostfeld says that Lyme disease should be tackled in part by targeting mice.

REALITY BITES

Not just an affliction of affluent New Englanders, tick-borne illnesses pose a threat to public health around the world. And they seem to be on the rise.

DISEASE	TICK VECTOR	GEOGRAPHY
>100,000 NEW CASES EACH YEAR		
Lyme disease *Borrelia burgdorferi* and other *Borrelia* bacteria	Blacklegged tick (*Ixodes scapularis*) and other *Ixodes* species	Global
10,000–100,000 NEW CASES EACH YEAR		
Tick-borne encephalitis (including Powassan and deer-tick virus) *Flavivirus* species	Various *Ixodes* species	North America, Europe, Russia, China, Mongolia
Spotted fevers (including Rocky Mountain spotted fever) *Rickettsia* bacteria	American dog tick (*Dermacentor variabilis*) and others	Global
1,000–10,000 NEW CASES EACH YEAR		
Babesiosis The protozoan *Babesia microti* and other *Babesia* species	Various *Ixodes* species	The Americas, Europe, southern and eastern Asia
Tick-borne relapsing fever *Borrelia miyamotoi* and other *Borrelia* species	Soft-bodied ticks of the genus *Ornithodoros*, and *Ixodes* species	The Americas, Europe, Asia, Africa
Ehrlichiosis and anaplasmosis *Ehrlichia* and *Anaplasma* bacteria	Lone star tick (*Amblyomma americanum*) and *Ixodes* species	Global
100–1,000 NEW CASES EACH YEAR		
Tick-borne tularaemia *Francisella tularensis* bacterium	Ticks of the genus *Dermacentor* and *Amblyomma*	Eastern and central United States, central Europe, Russia

I. scapularis
D. variabilis
O. hermsi
A. americanum

University in Richmond, says that he and his colleagues are working on an even better vaccine. One downside of an OspA vaccine is that it requires frequent boosters, because OspA antibodies have to be circulating constantly in the blood if they are to attack *B. burgdorferi* inside a biting tick. Marconi's team is developing a vaccine against immunologically relevant portions of the surface protein OspC, which *B. burgdorferi* expresses when it is inside mammals. On being bitten by infected ticks, vaccinated individuals can produce OspC antibodies from immunological memory; the antibodies do not have to be circulating already. Marconi and his colleagues have already licensed a version of the vaccine for use in dogs, and "the success of the canine vaccine and the uniqueness of the approach suggests that it's going to be highly effective in humans", he says.

In light of the problems faced by LYMErix, however, the question remains whether health officials and consumers will embrace a human vaccine. "I think, maybe optimistically, that the emotional situation has changed over the last 10 or 15 years — that is, that more people are convinced of the importance of Lyme disease," Plotkin says. But it is hard to know whether fears about Lyme will trump fears about the vaccine.

Mouse vaccines would not raise such concerns, but some researchers, including Plotkin, are sceptical about whether they could dose enough mice to reduce Lyme rates. And both vaccine approaches are limited because they combat only one tick-borne disease, when more than a dozen others are spreading throughout the world (see 'Reality bites').

TICK SPIT

There is one strategy that could conquer them all, and it involves turning one of the tick's most ingenious tools — its saliva — against it. When a tick bites a host, molecules in its saliva help it to evade detection and start to feed by blocking pain, inflammation and immune signals. If a vaccine could raise an immune response to key salivary proteins, it could make tick bites more noticeable or block the tick's ability to feed.

Ostfeld himself is a proof-of-concept for this approach. He has been bitten more than 100 times, and his body now reacts to tick saliva. "I realize when a tick is biting me because I get a burning sensation. It's pretty intense," he explains. Ostfeld has ample time to remove the tick before it can transfer an infection — if it even survives the experience. Often, Ostfeld says, he will remove a tick only to discover that, for unknown reasons, it is already dead.

A European Commission-funded consortium called ANTIDotE (Anti-tick Vaccines to Prevent Tick-borne Diseases in Europe) is characterizing the tick salivary proteins that could be targeted to thwart feeding. In 2011, a member of the group reported[6] a technique to rapidly identify those proteins that react with the blood serum of tick-immune animals. When the team vaccinated rabbits against three salivary proteins that it had identified — including one that ticks use to inhibit blood coagulation and one that inhibits the host's immune response — it found that ticks had trouble getting blood from them. Researchers in the group are also working to identify the salivary genes involved in *B. burgdorferi* transmission. "We think that an anti-tick vaccine could be immensely useful in protecting both humans and animals," says Hein Sprong, an ANTIDotE leader at the National Institute for Public Health and the Environment in Bilthoven, the Netherlands.

US Biologic also plans to develop a bait-based vaccine for mice that could thwart tick feeding, thereby protecting against multiple diseases. That could reduce overall tick numbers, too, because it would make it difficult for larval ticks to get the meals that they need to survive into adulthood and reproduce.

But these approaches are hardly around the corner. Part of the problem, scientists say, is that funding is scarce. The stereotype of Lyme and other US tick-borne diseases as primarily 'yuppie' illnesses does not help; Ostfeld says he has seen comments to this effect on reviews of his grant proposals. "They say something like, 'Is it really worth spending taxpayer dollars on a disease of the affluent in the northeastern United States, when there are so many diseases of people who live in poverty overseas?'," he says. "In one sense, I think that's a legitimate point, but in another, I think it underestimates the impact of this disease on a vast number of citizens, not all of whom are affluent, not even close." Another potential reason for low funding is that in the United States Lyme and similar infections are only rarely fatal. Each year, more people in the United States are diagnosed with Lyme than with prostate cancer, but research funding for the latter from the National Institutes of Health was more than ten times that for Lyme in 2014.

Until an all-encompassing solution becomes available, controlling tick-borne diseases will probably require an array of smaller-scale approaches that attack the problem, bit by bit, on a number of levels. That an arsenal of such weapons might be needed to hold back the enemy is not particularly surprising, considering the complexity of tick-borne-disease ecology, how drastically humans have been changing it, and how close people live to these disease-carrying parasites. "We've disrupted the balance of nature," Telford says. Steadying the scales again will be no small feat. ∎

Melinda Wenner Moyer *is a freelance science writer in Cold Spring, New York.*

1. Wilson, M. L., Telford, S. R. III, Piesman, J. & Spielman, A. *J. Med. Entomol.* **25**, 224–228 (1988).
2. Keesing, F. *et al. Proc. R. Soc. B* **276**, 3911–3919 (2009).
3. Richer, L. M. *et al. J. Infect. Dis.* **209**, 1972–1980 (2014).
4. Steere, A. C. *et al. N. Engl. J. Med.* **339**, 209–215 (1998).
5. Wressnigg, N. *et al. Lancet Infect. Dis.* **13**, 680–689 (2013).
6. Schuijt, T. J. *et al. PLoS ONE* **6**, e15926 (2011).

COMMENT

DAMON WINTER/NYT/REDUX/EYEVINE

Desert looms at the edge of Cathedral City, California.

Recognize anthropogenic drought

California's current extreme drought must be a lesson for managing water in a warmer, more densely populated world, say **Amir AghaKouchak** and colleagues.

Since 2012, California has been experiencing its worst drought in more than a century. Temperatures are breaking records and the region is down a year's worth of rainfall[1]. Forests, fish and wildlife as well as the regional economy are struggling.

California is known as the United States's salad bowl because of its prolific fruit, vegetable and nut production. But fields have had to be left fallow, contributing to statewide losses of US$2.2 billion in 2014 (ref. 2). More than 12 million trees have died (see go.nature. com/vrgp1e), with cascading impacts on amphibians, birds and mammals[3]. Streams and wetlands are drying up, including the American River hatcheries of steelhead and Chinook salmon. More than 17,000 jobs have been lost, mainly in poor rural communities[2].

California's ecosystems evolved under tremendous natural variation in weather and water availability. Extreme droughts tend to occur every century or so. But in the past 100 years, human activities have been drawing heavily on the water budget. Since 1950, growth in the state's population and agriculture have almost doubled water use. Whether California's drought is linked to the potential of rising greenhouse-gas emissions to increase the frequency of extreme weather is being debated[4]. The rise in ▶

water demands is not.

Overuse and obsolete management of scarce water resources are exacerbating the current drought's impacts. Past leaders legislated for and invested in measures and infrastructure to boost supplies as demand grew. Now the state is nearing its water limits and can no longer simply build its way out.

California's water troubles are a harbinger of things to come around the world, wherever population and industries are growing. We call for more studies and legislative consideration of the human impacts on water stress caused by urbanization, greenhouse-gas emissions and food and energy production, as well as for policy and management practices more suitable to prosperous economies and developed water systems.

ADAPTATION LEGACY

California has a long history of adapting to droughts. Droughts in the nineteenth century struck an economy then based on free-ranging cattle and rain-fed agriculture that demanded more irrigation[5]. Population growth in the early twentieth century led to the building of major reservoir, canal and hydropower networks across the state.

More recent droughts, in 1976–77, 1988–92 and 2007–09, led to aggressive urban conservation programmes and greater use of water markets, such as the Emergency Drought Water Banks of the 1990s that facilitated the purchase, sale and transfer of water. Other responses included increased groundwater extraction and use, water conservation, more-extensive irrigation and more infrastructure for conveyance, storage, wastewater re-use and brackish-water desalination.

Most of these innovations have served to maintain or expand water supplies. Today, the state has one of the world's most engineered and diversified water systems, including six aqueduct networks more than 2,000 kilometres long and in excess of 1,400 dams. Despite the current drought's severity, urban water supplies have been managed and had few major interruptions.

"The pinch has mainly been felt in rural communities and in the environment."

The pinch has mainly been felt in rural communities and in the environment. Because California's land-use legacies — of rapid urbanization, wetland and floodplain development and more-dense forests — have exacerbated the environmental impacts of drought, over the past century, natural habitats have shrunk and been deprived of water diverted to meet human needs. California still lacks comprehensive drought management plans for land, rivers and wetlands. Little policy addresses how upland changes affect water balance across a catchment.

California has one of the world's most engineered water systems, including six aqueduct networks.

Increased groundwater extraction lowers the base flows of rivers, streams and pools. Pool networks become disconnected and food webs are broken[6]. Intermittent streams, which provide rearing and breeding habitats for river biota, are especially vulnerable to drought[7]. Reduced stream flow and snow melt in warm dry periods[8] further diminish natural groundwater and stream flow in a cycle that compromises habitats for native fish, including delta smelt and salmon[6,8].

The reductions in water diversions in 2012–13 to protect endangered species in the Sacramento–San Joaquin River Delta region have meant less water for urban and agricultural users. Curtailments in surface-water pumping to protect endangered species limit supplies to the Central Valley for agriculture, increasing unemployment among farm workers. Political conflict over urban, environment and agricultural water rights has erupted.

The current drought has already led to the state's most significant water initiative in half a century. The Sustainable Groundwater Management Act of 2014 establishes a framework for long-term groundwater monitoring, assessment and management statewide. It tasks local agencies with assessing groundwater conditions in basins and developing local water management, allocation and adaptation plans. This is a major breakthrough for sustaining groundwater, although full implementation of the act is expected to take decades.

In July, the US House of Representatives passed legislation that, if it becomes law, will offer some drought relief to Californian farmers and growers at the cost of protecting endangered fish. The legislation would take more water from the Sacramento–San Joaquin River Delta for agricultural users, and away from rivers. Such dramatic policy responses may have irreversible impacts, such as the extinction of native fishes.

CLIMATE CHANGE

Severe and long-lasting droughts have occurred in reconstructions of the region's past climate[9]. So it is not clear whether California's current drought is a temporary weather condition or is the emergence of a 'new normal'. Observations and climate projections indicate that California's climate is warming, with more winter rainfall instead of snow, earlier snow melt and decreases in spring and summer stream flows[8].

Future droughts will be compounded by more-intense heatwaves and more wildfires. Soaring temperatures will increase demand for energy just when water for power generation and cooling is in short supply. Such changes will increase the tension between human priorities and nature's share.

As anthropogenic effects increase, lessons from past droughts cannot simply be applied to future events. More research is needed on the potentially different water supply and demand that will characterize droughts in a warmer world. California's resilience to such future events has not been examined. Climate and hydrological scientists focus on large-scale phenomena and give little attention to local conditions and impacts, such as reduced economic production or depletion of local groundwater.

NEXT STEPS

California must learn to live with its dry climate. Rather than responding to crises, the state needs a proactive and long-term drought-management plan that considers all sectors, including the environment.

PETE MCBRIDE/NATL GEOGRAPHIC CREATIVE

Left: More than 12 million trees have died from California's drought. Right: Irrigation manager Rick Gilmore inspects dried grape vines in Byron, California.

Regional adjustments to reduced precipitation or snow pack are needed, even if overall precipitation does not change. Demand management, conservation, public outreach, technological innovation for water conservation and more-flexible market-based solutions and infrastructure adaptation are fundamental to responding to increased demands and climate-change stress in the future.

For example, Australia's 1997–2009 Millennium Drought triggered changes in public perception and policy reforms, water demand and other environmental management strategies[10]. City dwellers reduced their water use. Incentives for using water-conservation technologies slashed long-term water demand.

Breakthroughs in drought management, adaptation, mitigation, resilience assessment and prediction demand the close collaboration of scientists, policymakers and decision-makers. Federal and state agencies should develop long-term research programmes to address key science challenges and the realization of innovative technologies.

Explorations of drought must go beyond supply issues to encompass water demands (including environmental needs), water-storage infrastructure, adaptability, policy and feedback between human actions and climate as a complex system. This requires embracing a broader definition of socio-economic drought — a condition in which water demands exceed available supply.

A key element is investing in drought-monitoring and -prediction systems. Over the past decades, drought monitoring has evolved to include satellite observations and advanced analysis software. But drought prediction at seasonal scales remains a challenge. Climate and hydrological scientists should prioritize such research to support better water management and explore the causes, local impacts and management of droughts across many scales. The social consequences of downsizing agriculture during water-scarce periods require deeper research and planning.

Decision-makers should update drought-preparation strategies given that combined demand growth and reduced water supplies make it more difficult to recover after droughts. This will entail prioritizing water needs under extreme conditions in a way that minimizes long-term impacts on both humans and ecosystems. A range of behavioural and technological options (such as conservation and water reuse) must be considered. Water-policy reforms and the establishment of water entitlements for environmental protection during the Millennium Drought are examples of successful adaptation plans.

Policymakers should establish environmental water entitlements and drought plans, based on understanding water needs for ecosystems and trade-offs between endangered species and crucial water uses. This should include emergency actions for key river segments and refuge habitats, including evacuation and captive breeding to avoid extinction of endangered species.

Water and environment managers must reconcile environmental water supplies with economic water uses, and develop adaptive plans for future conditions. Long-term impacts of policy reforms should be considered to avoid further socio-economic impacts. Industry and the agricultural sector should prepare for a warmer climate and lessen their water use by increasing water recycling and efficiency and forgoing low-value water uses.

California's current extreme drought must be a lesson for managing water in a warmer future climate with increased demands. ∎

Amir AghaKouchak *is an assistant professor in the Center for Hydrology and Remote Sensing, University of California, Irvine, USA.* **David Feldman** *is a professor in the Department of Planning, Policy and Design, University of California, Irvine, USA.* **Martin Hoerling** *is a meteorologist at the National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Boulder, Colorado, USA.* **Travis Huxman** *is a professor in the Department of Ecology and Evolutionary Biology, University of California, Irvine, USA.* **Jay Lund** *is a professor in the Department of Civil and Environmental Engineering, University of California, Davis, USA.*
e-mail: amir.a@uci.edu

1. Shukla, S., Safeeq, M., AghaKouchak, A., Guan, K. & Funk, C. *Geophys. Res. Lett.* **42,** 4384–4393 (2015).
2. Howitt, R. E., Medellin-Azuara, J., MacEwan, D., Lund, J. R. & Sumner, D. A. *Economic Analysis of the 2014 Drought for California Agriculture* (University of California, Davis, 2014).
3. Adams, H. D. *et al. Proc. Natl Acad. Sci. USA* **106,** 7063–7066 (2009).
4. Herring, S. C., Hoerling, M. P., Peterson, T. C. & Stott, P. A. (eds) *Bull. Am. Meteor. Soc.* **95,** S1–S96 (2014).
5. Lund, J. R. & Medellin-Azuara, J. in *Proc. World Environmental and Water Resources Congr.* 2073–2080 (American Society of Civil Engineers, 2015).
6. Power, M. E., Bouma-Gregson, K., Higgins, P. & Carlson, S. M. *Copeia* **103,** 200–211 (2015).
7. Hwan, J. L. & Carlson, S. M. *River Res. Applic.* http://dx.doi.org/10.1002/rra.2907 (2015).
8. Barnett, T. P. *et al. Science* **319,** 1080–1083 (2008).
9. Diaz, H. F. & Wahl, E. R. *J. Climate* **28,** 4637–4652 (2015).
10. AghaKouchak, A. *et al. Science* **343,** 1430–1431 (2014).

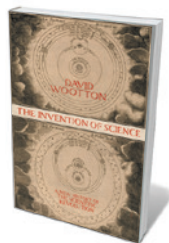HISTORY OF SCIENCE

The crucible of change

Philip Ball gets to grips with a revolutionary history of the scientific revolution.

David Wootton's *The Invention of Science* will cause arguments, and is all the better for it. The idea of a 'scientific revolution' — the supposed birth of modern science, beginning around the start of the seventeenth century and emerging from the work of Galileo Galilei, Johannes Kepler, Isaac Newton, Robert Boyle and their contemporaries — has been rejected by many historians of science. But Wootton argues that the period marked a true turning point, in which the whole process of doing science was transformed into a social enterprise with established norms and procedures.

Nay-sayers deny that any 'essence' of modern science crystallized in this period. As science historian Steven Shapin wrote at the beginning of his 1996 *The Scientific Revolution* (University of Chicago Press), "There was no such thing" as this revolution. But Wootton uses the term unapologetically, and assesses the contributions of the luminaries of the period from a stance that he admits is Whiggish — essentially, judging their ideas in the light of present understanding.

All too often, such a "presentist" view is taken to extremes by scientists writing history, either because they don't know any better or because, like physicist Steven Weinberg in *To Explain the World* (Allen Lane, 2015), they believe that the history of science belongs to its practitioners (see R. P. Crease *Nature* **518**, 300; 2015). This manifests as a tendency to dish out medals for getting things 'right' — leading Weinberg to dismiss Plato as silly, philosopher Francis Bacon as irrelevant and Galileo's errors as the result of off-days. In this view, pursuits such as alchemy and astrology were useless mumbo jumbo, and Newton and Boyle let the side down by practising them. Finding out about nature was just a matter of doing the right experiments and interpreting them with the proper scientific method.

Wootton's book is very different, however. It is perceptive, thought-provoking, deeply erudite and beautifully written. Wootton argues that

The Invention of Science: A New History of the Scientific Revolution
DAVID WOOTTON
Allen Lane: 2015.
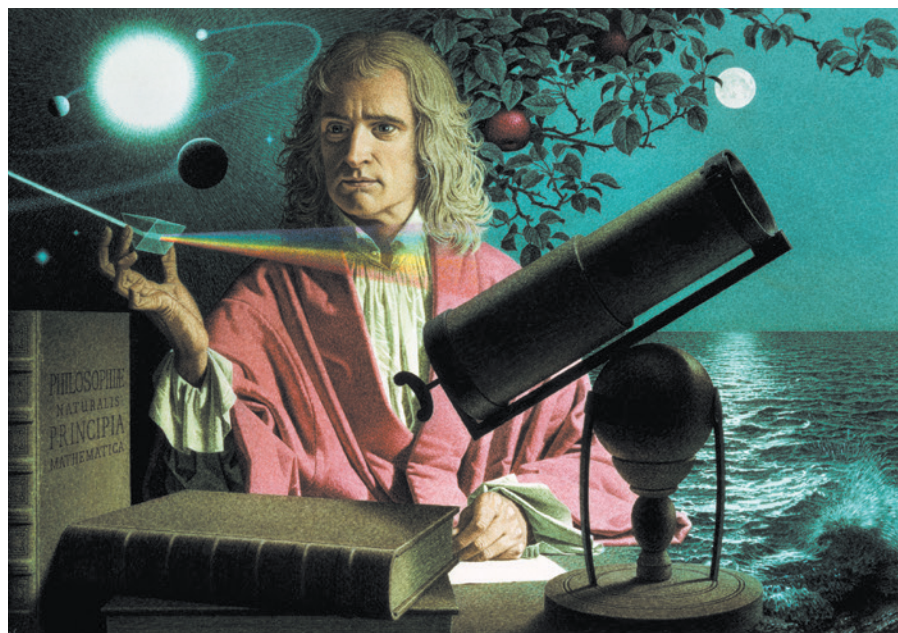
Isaac Newton, drawn in 1974 by Jean-Léon Huens, was at the heart of the scientific revolution.

the scientific revolution took place between 1572 and 1704, and contends that it led to genuine intellectual progress and to a new, more effective way of understanding the world. Does he make a persuasive case? I'll come back to that.

Those two dates hint at what Wootton deems crucial. Danish astronomer Tycho Brahe's observation of a nova (which he interpreted as a new star) in 1572 fatally undermined the Aristotelian concept of the unchanging heavens. The publication of Newton's *Opticks* in 1704 completed the Newtonian revolution, rendering the world mechanical and mathematical. The crucible of change was thus the seventeenth century: the time when Bacon argued for a systematic approach to what we now call science, Galileo asserted that mathematics is the language of nature, and institutions including the Royal Society in London formalized how science was studied. At the outset, the very meanings of concepts such as experimentation were unclear, and classical authorities such as Aristotle still held sway. By the end, a recognizably modern science existed.

That much is broadly accepted, and many of the milestones Wootton covers are familiar. These include the exploration of the geography, flora and fauna of the New World and its relevance to debates about the structure of the (spherical) Earth; the invention of the telescope and Galileo's observations of the Moon and the satellites of Jupiter; and Newton's 1687 book *Principia Mathematica*.

What elevates Wootton's account is how he uses a close and perceptive reading of such developments to explore changes in the way nature was investigated and conceptualized. He argues that the transformation was in how knowledge was obtained, recorded and assessed, even more than in what was known about the world.

The core of the book is conceptual, not chronological: Wootton looks at changing meanings of words such as fact, experiment, theory and evidence. 'Fact', for instance, was originally a legal term meaning a deed or act. It was not until the 1660s that it was used in England to signify a reproducible piece of knowledge about the world. Wootton argues that this usage became inevitable with the emergence of a print culture that, by an almost Darwinian process, let knowledge replicate fast enough to drive out error with an accumulation of evidence.

It is precisely because of this attention to context that *The Invention of Science* is not as heretical as it thinks it is. Wootton's tilting is against the 'relativist' position that provoked the 'science wars' of the 1990s, in which scientists including Weinberg were dismayed by historians who argued that social factors such as authority and status shaped how

science evolved. At their worst, relativists maintained that scientific knowledge was entirely a social construct, no more 'true' than Galenic medicine or witchcraft. It is because this absurd argument still persists among some historians that Wootton's nuanced approach is needed.

But few historians of science are that extreme, and Wootton sometimes exaggerates the position of the milder 'relativists' who he says dominate the discipline today. For example, he takes particular issue with Shapin's highly influential 1985 collaboration with science historian Simon Schaffer, *Leviathan and the Air-Pump* (Princeton University Press), which argued that the reception of Boyle's experiments on gases and vacuums had as much to do with politics as with objective 'evidence'. But Shapin and Schaffer do not imply that Boyle's gas law is, as a result, mere convention. Rather, they challenge the naivety of the idea that science, proceeding openly and aloof from its sociopolitical environment, reaches incontrovertible truths by unassailable reason. Wootton's equally valid point is that the undoubted contingency of science's methods and context does not make its results any less reliable, or not always. The two views do not seem incompatible.

"What marks out modern science", Wootton writes, "is not the conduct of experiments … but the formation of a critical community capable of assessing discoveries and replicating results". Science needed to be reported openly and debated by peers, as it was (after a fashion) in the Royal Society's *Philosophical Transactions*, the first true scientific journal, launched in 1665. It is a very plausible case, and one that most historians of science should not find too hard to swallow.

> **"Wootton comes not to bury relativism, but to curb its excesses."**

In any event, Wootton admits that he is seeking a middle ground: he comes not to bury relativism, but to curb its excesses. Far from renewing hostilities, this timely and thoughtful book should encourage historians of science to discover how much they agree on, and to refine the points of dispute. "The task," Wootton says, "is to understand how reliable knowledge and scientific progress can and do result from a flawed, profoundly contingent, culturally relative, all-too-human process." That is beautifully put and, in my view, right on the mark. ∎

Philip Ball *is a writer based in London.*
e-mail: p.ball@btinternet.com

Subverting surveillance

Anthony King tours a playful exhibition that probes covert data collection and tracking.

Transparency Grenade by Julian Oliver explores the ease of leaking information.

When US computer analyst and government contractor Edward Snowden began lifting the lid on the covert surveillance activities of the US National Security Agency (NSA) in 2013, the social implications shocked even cynics. Now, Secret, the latest show at Dublin's Science Gallery, probes this dark side of the digital age through 23 eclectic exhibits.

Curated by Ian Brunswick, Marie Redmond, Tad Hirsch and Julian Oliver, Secret asks what is being hidden from us, what we are hiding and why. It pushes visitors to debate the automated collection of data by businesses and governments. And it prompts us to playfully subvert mass surveillance, echoing the gallery's 2012 exhibition Hack the City (see *Nature* **486**, 470; 2012) and its call to take back urban spaces and services.

Once inside the gallery, first up is *Please Empty Your Pockets* by artist Rafael Lozano-Hemmer — an exhibit inviting you to decant your possessions onto a conveyor belt with a scanner. It feels invasive as you divest yourself of your smartphone. And it reminds you that such 'exposure' is nothing compared to how our movements, conversations and personal data are collected, archived and utilized by others every day, often by means of smartphones set to track our every move.

You can fight back — for instance, throwing a spanner in the works of Big Brother by using artist Benjamin Grosser's *ScareMail*. This browser extension supplements e-mail signatures with spurious algorithm-generated narratives, replete with probable NSA search terms such as 'facility', 'plot' or 'dirty bomb'. E-mail-surveillance programs aim to detect certain keywords, so such bogus stories disarm them. Another browser add-on, *AdNauseum* by artist Daniel Howe and designer Mushon Zer-Aviv, blocks ads on web pages while simultaneously clicking each ad in the background — dual action that whitewashes a user's profile with a meaningless spray of data. Its anarchic aim is to fan mistrust between advertisers and ad networks.

Downstairs, you get a chance to sample the fruit of a bizarre digital accident. Artist Zoë Irvine unwittingly recorded two years' worth of her phone calls — but only her side of the conversation. In *Your Voice Is You*, you can listen in on a white rotary-dial telephone. Using voice-analysis

Secret: Nothing to See Here
Science Gallery, Trinity College Dublin.
Until 1 November.

software, Irvine has organized her one-sided chats under the categories truth, confusion, embarrassment, stress, excitement, anticipation and suspected lies. It feels oddly intrusive, even given Irvine's licence to eavesdrop.

Visitors can also play intelligence analysts in the interactive artwork *Crowd-Sourced Intelligence Agency* by Jennifer Gradecki and Derek Curry. Faced with 100 tweets from each of more than 26,000 Science Gallery Twitter followers, you are asked to tag a post as threatening, non-threatening or flagged for review by other agents, adding notes to explain your decision. "It opens up the process for people to participate in surveillance themselves," explains Gradecki. When a tweet is evaluated, the original poster is notified and given a link to the post on the artwork's website. The point? That such overt, democratic involvement is denied us when government or corporate entities explore social media.

There is a frolicsome flavour to the *Spy Puzzle Café* by artist Ruben Pater (working as Untold Stories), where you can test your puzzle-solving acumen by navigating a maze in the form of a portrait of Snowden. And there are special events: for example, you can try your hand at lock-picking in a two-hour workshop for the true enthusiast. Lock-picker and ethical hacker Martin Mitchell will reveal how to unlock devices by analysing and manipulating the components. Of course, it is what you do with this knowledge that matters; as exhibition researcher Zack Denfeld puts it: "We want to encourage impish curiosity."

This is a show that strives to teach, tease and irk visitors into considering the ubiquity of secrets and secret-breakers. There is enough thought-provoking fun here to deflect the worst of any angst over digital surveillance, even as it continues to nibble at our collective psyche. ∎

Anthony King *is a writer based in Dublin.*
e-mail: anthonyjking@gmail.com

SCIENCE FICTION

Cosmology boot camp

John Gilbey goes on the road in the US far west to refine the science in his fiction.

I was looking for a man with a small dragon perched on his shoulder, but the low atmospheric pressure made even the short walk from the shuttle to the concourse more taxing than I had expected. Above me, the building opened up into a vast, vaulted tent of brilliant white fabric. Beyond the tinted windows, the arid landscape looked almost abstract, a glint of ice visible on the peaks of the distant mountains.

The scene was familiar from a dozen science-fiction novels, but this was no poorly pressurized dome on some airless future world. It was Denver International Airport in Colorado — a mile above sea level on the edge of the Rocky Mountains — and the dragon was a detailed model sported by one of my fellow science-fiction creators. We were meeting there to head northwards together to the Launch Pad Astronomy Workshop at the University of Wyoming in Laramie.

Launch Pad is the creation of Mike Brotherton: astronomer, sci-fi novelist and professor in the physics and astronomy department at Wyoming. Now in its ninth year, the programme is a wildly enjoyable week-long boot camp for science-fiction practitioners determined to get the astronomy and cosmology right in their stories, films, graphics and games. The list of authors associated with Launch Pad as instructors and students includes the likes of Joe Haldeman and Mary Robinette Kowal. They represent a wide range of sub-genres, from hard science fiction to fantasy, and explore subjects from black holes to werewolves.

Laramie — at an altitude of more than 2,100 metres and with a dry, stable climate — is an ideal place to combine classroom and lab activities such as learning the art and science of spectroscopy and Hubble image management with hands-on experience of a range of telescopes. Observing Jupiter's moons as the sky darkened above the Rocky Mountains is an experience that will stay with me for a long time. A visit to the impressive 2.3-metre telescope at the Wyoming InfraRed Observatory is a pivotal part of the course and a brilliant opportunity for the team photo.

After a week grappling with cosmological theory — not the least mind-mangling part of which was dark matter — and just geeking out with Mike, his colleagues and
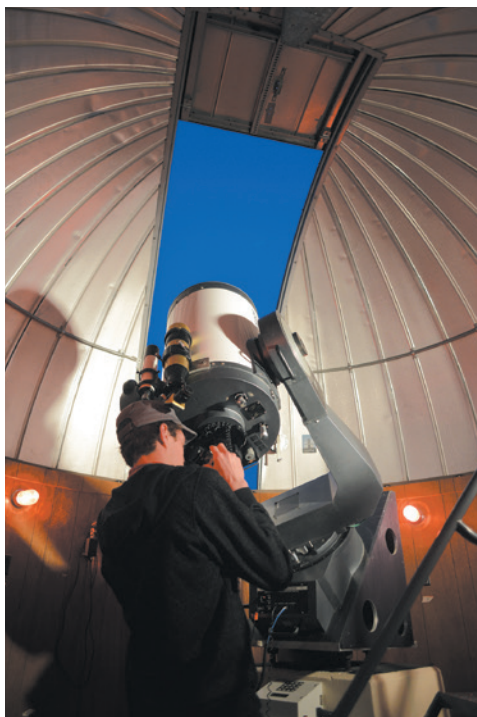
Launch Pad assistant Tyler Ellis sets up a telescope.

other writers, the workshop came to an end. Alone, I headed west by train through the Rocky Mountains — a welcome digital detox after immersion in technology, through glorious, almost Martian scenery. Thirty-three hours later I was decanted into the intellectual maelstrom of California's Silicon Valley, keen to discover what the wider science community thinks of science fiction.

A scale model of the Fermi Gamma-ray Space Telescope hangs in the foyer of the Kavli Institute for Particle Astrophysics and Cosmology (KIPAC) on the sandstone hills above Stanford University. It is framed by translucent marker boards covered in the arcane glyphs of the trade. This space and the neighbouring SLAC National Accelerator Laboratory could easily serve as the set of a hi-tech sci-fi blockbuster. KIPAC director Tom Abel spent time with me describing the multi-threaded nature of the institute's work in the search for dark matter and dark energy. Parallel streams of research with a common goal are rarely seen in science fiction — especially in films, where science is frequently simplified.

Kludgy science in fiction does not bother Abel. "I actually practise suspending my criticism as much as possible in order to maximize the enjoyment," he told me with a smile. Stories of future human society appeal to him. "I keep looking for the philosophical, the sociological, the human dimensions — there are so many challenges there, balancing individual fulfilment with a million-year timescale." I began wondering whether hard-science researchers prefer social-science fiction, and decided that further research would be needed.

A couple of miles away, on a busy street corner in Palo Alto, the non-profit research organization Institute for the Future treads an intriguing line between science futures and science fiction in providing "practical foresight for a world undergoing rapid change". Mike Liebhold, a distinguished fellow at the institute, is in the forecasting business. As he tells me, "One of the most productive methods we use is provocation, to help people take a much narrower focus than merely plausible futures, and actively engage in shaping their desirable futures." He walked me around the centre's frenetic creative space. "I'm trying to understand how humans can orchestrate automation and digital technology," he notes — a theme common to many a sci-fi plot.

As a writer in my late fifties, I am frequently tripped up by the sense of living in my past-future, surrounded by both the shiny futuristic gadgets and the dark dystopias that my younger self read and wrote about. (Compare, for instance, the Newspad in Stanley Kubrick's 1968 film *2001: A Space Odyssey* and the handheld communicator in the original series of *Star Trek* with today's tablets and smartphones.) Now, more than ever, I believe that the project management of our collective future needs to work across science cultures to gain from all the vitally creative contributions available. By fostering the inclusion of sound science in future fiction, events such as Launch Pad help to build mutual awareness between science practitioners and science speculators. This will help us to work towards the goal of a secure future for our species and home planet. ■

John Gilbey is a UK-based science and science-fiction writer. He teaches computer science at Aberystwyth University.
e-mail: gilbey@bcs.org.uk

Correspondence

Cut food waste to help feed world

Climate-proofing farms to help feed the world's expanding population needs to be complemented by global measures to cut food losses and waste (see *Nature* **523,** 396–397; 2015).

Roughly one-third of the food produced annually for human consumption — around 1.3 billion tonnes per year — is lost or wasted worldwide (see M. M. Rutten *Agric. Food Secur.* **2,** 13; 2013). In developing economies, more than 40% of food losses occur after harvesting or during processing (see go.nature.com/um7vga).

We should follow the lead of the Asia-Pacific economies, which have pledged to reduce food losses and waste by 10% relative to 2011–12 levels by 2020 (see go.nature.com/rfupkz).
Don Gunasekera *Victoria University, Melbourne, Australia.*
don.gunasekera@vu.edu.au

Forgotten findings of brain lymphatics

Antoine Louveau and colleagues describe lymphatic vessels in the central nervous system (*Nature* **523,** 337–341; 2015), suggesting that "the unique location of these vessels may have impeded their discovery to date". However, these findings are not without precedent.

The first description of a lymphatic circulation in the head was soon forgotten (G. Schwalbe *Z. Med. Wiss.* **7,** 465–467; 1869), as were other early landmark experiments (J. B. Brierley and E. J. Field *J. Anat.* **82,** 153–166; 1948). These included half a dozen seminal contributions in the 1960s from a Hungarian group led by the late neurologist Endre Csanda and lymphologist Mihaly Földi.

This group of researchers described the existence and significance of a lymphatic system in the central nervous system,
stating that "in striking contrast to the textbook opinion, lymph drainage plays an important part in fluid circulation of the brain". They identified "well-defined connections between the subarachnoid space and the cervical lymph system in the nasal cavity, in the orbita and in the region of the jugular foramen" (M. Földi *et al. Angiologica* **5,** 250–262; 1968).

Perhaps old papers that are not available as online PDFs are easily overlooked in today's literature searches, depriving many of our scientific predecessors of the recognition they deserve.
Éva Mezey *National Institute of Dental and Craniofacial Research, Bethesda, Maryland, USA.*
Miklós Palkovits *Semmelweis University, Budapest, Hungary.*
mezeye@nidcr.nih.gov

Lifelong learning for all in biomedicine

Cross-disciplinary and team-based modern research is overwhelming established mechanisms for maintaining professional competency. This calls for a change to personnel training that is not limited to professors (see C. E. Leiserson and C. McVinney *Nature* **523,** 279–281; 2015).

The European Innovative Medicines Initiative projects on education and training aim to promote lifelong learning and continuing professional development among all biomedical professionals (see go.nature.com/enhijn). We are working with employers, professional and scientific bodies, course providers and individual scientists to develop a framework for implementing this cultural change, known as LifeTrain (see M. Hardman *et al. Nature Rev. Drug Disc.* **12,** 407–408; 2013).

By providing guidance and information, we help them to recognize the competencies that they will need as their careers progress. We also encourage
scientists to think beyond traditional disciplinary silos.

Several organizations and individuals have formally signed up to the agreed principles and implementation of LifeTrain; other prospective participants can contact us at www.lifetrain.eu.
Cath Brooksbank, Claire Johnson *EMBL-European Bioinformatics Institute, Hinxton, Cambridge, UK.*
cath@ebi.ac.uk

Physicist's death changed war policy

Just over 100 years ago, on 10 August, the 27-year-old British physicist Henry Moseley was killed in the First World War at the battle of Gallipoli. His work on the X-ray spectra of atoms had already explained the basis of Dmitri Mendeleev's periodic table of the elements, laying the foundations for the interpretation of chemistry through atomic structure.

Ernest Rutherford, who discovered the atomic nucleus, wrote an obituary of Moseley (*Nature* **96,** 33–34; 1915) that may have helped to permanently change government policy on enlisting prominent scientists for combat duty. Even in the heat of war, Germany, too, lamented Moseley's death, such was his contribution to science (K. Fajans *Die Naturwissenschaften* **4,** 381–382; 1916).
Min-Liang Wong *National Chung-Hsing University, Taiwan.*
mlwong@dragon.nchu.edu.tw

Offsets: conservation served by flexibility

The debate over whether national protected areas are eligible for biodiversity-offset funding should factor in the different challenges and contexts for countries seeking to conserve their biodiversity (see M. Maron *et al. Nature* **523,** 401–403; 2015).

Offsets that are voluntary (such as those of the mining
companies Minera Panamá and QIT Madagascar Minerals; see also go.nature.com/zxv1et) are a source of extra conservation funding. This is analogous to grants from foundations or from the Global Environment Facility, which are considered a legitimate way to finance fulfilment of the Aichi biodiversity targets (see go.nature.com/avssmr).

In countries where offsets or compensation payments are required by law or lender standards, they serve as a tax on environmental impacts. This follows the accepted 'polluter pays' taxation model, and has the benefit of earmarking the revenue for conservation.

In heavily populated regions, the only socially acceptable place to implement offsets might be in protected areas selected by public processes. In such settings, offsets and protected areas would need to be closely linked.

Excluding offsets or compensation from financing of protected areas would affect developing countries, most of which struggle to finance such areas. It would, for example, exclude all of Brazil's protected-areas system from contributing to Aichi Target 11, because compensation is integral to the country's law on protected areas.
Jared J. Hardner *Hardner & Gullison Associates, Amherst, New Hampshire, USA.*
Raymond E. Gullison *Hardner & Gullison Associates, Lantzville, British Columbia, Canada; and University of British Columbia, Vancouver, Canada.*
Porter P. Lowry II *Missouri Botanical Garden, St Louis, Missouri, USA.*
ted@hg-llc.com

Yoichiro Nambu
(1921–2015)

Visionary theorist who shaped modern particle physics.

Yoichiro Nambu was one of the most influential theoretical physicists of the twentieth century. His deep and unexpected insights often took years for others to understand and fully appreciate. They include: spontaneous symmetry breaking, for which he was awarded half of the 2008 Nobel Prize in Physics; the theory of quarks and gluons; and string theory.
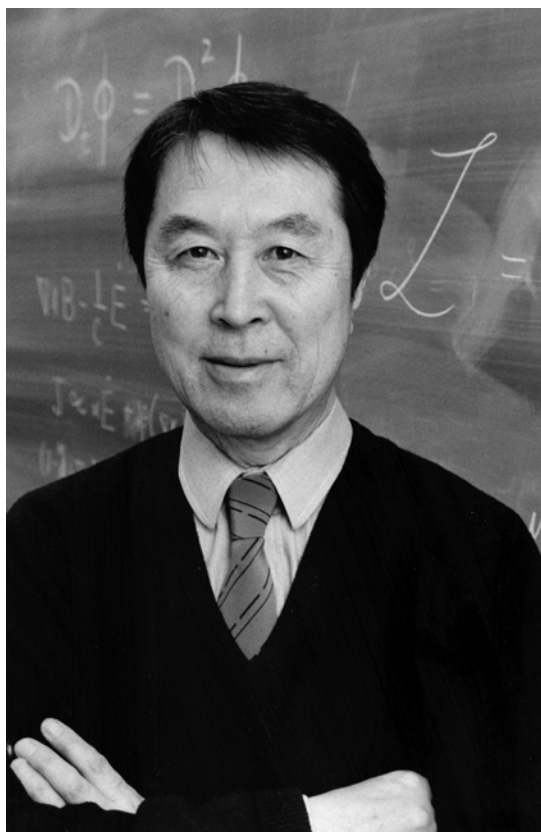
Modern particle theory is defined by its accomplishments, largely embodied in the standard model of the strong, weak and electromagnetic interactions, and by its aspirations — a theory that unifies all the forces and particles. Nambu's contributions to symmetry breaking and the theory of quarks form the foundation of the standard model, and string theory is the most promising approach to a theory of everything.

Nambu, who died of a heart attack on 5 July in Osaka, Japan, was born in Tokyo in 1921. It was the year that Yoshio Nishina visited Copenhagen and brought back quantum mechanics to Kyoto, Japan's first foray into modern physics. The 'Copenhagen in Kyoto' school included Hideki Yukawa, who won the 1949 Nobel prize for his prediction of the existence of mesons, and Sin-Itiro Tomonaga, who shared the 1965 Nobel prize for his work in quantum electrodynamics, the theory that describes all of electromagnetism from chemistry to lasers.

Nambu attended the University of Tokyo, graduating with a master's degree in physics in 1942. His studies were interrupted by the Second World War. In the army he dug trenches and worked on the Japanese radar project, but his mind was on fundamental physics. In 1945, he married his assistant, Chieko Hida.

Under difficult post-war circumstances, always hungry and living in his office at the University of Tokyo, Nambu finished his PhD in 1952. Although his department's research focus was condensed-matter physics, Nambu was drawn to nuclear and particle physics and he attended seminars on these topics by Nishina, Tomonaga and Yukawa at the nearby Tokyo University of Education.

In 1950, Tomonaga recommended Nambu for a faculty position at Osaka City University, where he wrote two remarkable papers. He derived the now-famous Bethe–Salpeter equation that describes the quantum theory of how particles bind together (Y. Nambu *Prog. Theor. Phys.* **5,** 614–633; 1950). And he proposed how the newly discovered 'strange' particles were produced (Y. Nambu *et al. Prog. Theor. Phys.* **6,** 615–619; 1951). Each paper pre-dated by a year its more well-known counterpart written by US physicists.

Nambu's big break came in 1952 when, at Tomonaga's suggestion, he was invited by Robert Oppenheimer to the Institute for Advanced Study in Princeton, New Jersey. Years later he described that experience as overwhelming — he felt surrounded by people smarter and more aggressive than him. Nonetheless, physicist Murph Goldberger thought highly enough of him to invite him to the University of Chicago, Illinois, in 1954.

In particle physics, Chicago was the place to be just after the Second World War. Enrico Fermi was the intellectual leader of a physics department that included more than ten future Nobel prizewinners. Nambu spent the rest of his academic career — more than half a century — at the university's Enrico Fermi Institute.

With Giovanni Jona-Lasinio in 1961, Nambu introduced the idea of hidden or broken symmetries while trying to understand superconductivity — the resistanceless flow of electric current at very low temperatures. Mathematical symmetries in Maxwell's theory of electromagnetism are hidden at very low temperatures, as is the symmetry between the electromagnetic and weak forces, the hallmark of the electroweak theory. The Higgs boson, discovered in 2012 at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, reveals the fact that the electroweak symmetry is broken.

In 1964, George Zweig and Murray Gell-Mann each independently proposed the idea of quarks to explain the hundreds of new elementary particles that were being discovered at particle accelerators. It took more than 20 years to sort out quarks' properties and how they are held together in triplets and pairs by a 'colour' force mediated by gluons to form protons, neutrons, mesons and other particles. But Nambu and Moo-Young Han put most of it together in 1965. As Gell-Mann said: "He did this … while the rest of us were floundering." In an attempt to better understand the colour force, Nambu went on to co-invent string theory.

I had the good fortune of being Yoichiro's colleague for more than 30 years. He was surprisingly soft-spoken and modest for someone so wise and important. We all listened carefully to anything he had to say, but rarely fully comprehended it. "People don't understand him, because he is so far-sighted," Edward Witten of the Institute for Advanced Study once said.

A downside of being so ahead of the times is that recognitions come slowly. After years of hoping, we were ecstatic when he received the Nobel prize. Nambu could not travel to the ceremony in Stockholm, so the Swedish ambassador to the United States came to Chicago to present the modest giant of particle physics with his prize at ceremony attended by 200 of his friends and colleagues. A more joyful event I cannot remember. ■

Michael S. Turner *is professor of astronomy and astrophysics, and of physics, and director of the Kavli Institute for Cosmological Physics at the University of Chicago, Illinois, USA.*
e-mail: mturner@kicp.uchicago.edu

MATERIALS SCIENCE

Superlattice substitution

What happens if some of the particles of a superlattice — an array of identical nanoscale crystals — are replaced with foreign ones? It emerges that the properties of superlattices can be radically altered in this way. SEE LETTER P.450

DANIEL VANMAEKELBERGH

For two decades, researchers have been preparing solid materials that consist of ordered arrays of nanometre-scale particles. By analogy with atomic lattices, the incorporation of traces of particular nanocrystals into these superlattices might allow the materials' properties to be tailored. On page 450 of this issue, Cargnello et al.[1] report the formation of two-dimensional and thin-film superlattices of cadmium selenide (CdSe) or lead selenide (PbSe) nanocrystals that have been 'substitutionally doped' with nanoparticles of gold or gold–silver alloy — that is, controlled amounts of these nanoparticles have been incorporated so that they occupy lattice sites in the host material. The authors find that this does indeed alter the superlattices' properties in potentially useful ways.

The incorporation of foreign atoms into a host material's atomic lattice is an essential step in the fabrication of materials for communications technology, opto-electronic devices and construction. A classic example is the introduction of arsenic or phosphorus into a silicon crystal — each doped atom forms four bonds with neighbouring silicon atoms and donates its remaining valence electron to silicon's conduction band. This transforms silicon from an electrical insulator to the most widely used semiconductor. But the effects of analogous substitutional doping in

superlattices have not been explored.

Nanocrystal superlattices emerged not long after the discovery[2] that chemical synthesis could be used to produce suspensions of nanocrystals in solvents, in which the nanocrystals all have the same size and shape. Superlattices can form from these suspensions through crystallization induced by evaporation of the solvent under controlled conditions[3]. In their work, Cargnello and co-workers mixed a suspension of CdSe (or PbSe) with gold nanocrystals in hexane, and cast it on an immiscible liquid (ethylene glycol), which acted as a substrate on which superlattices could form. In this way, they prepared monolayers, bilayers and thin films of nanocrystals as ordered arrays, which exhibited astonishing regularity over hundreds to thousands of unit cells (the smallest units of a crystal lattice).

Using a careful statistical analysis, the authors demonstrated that dopant gold nanocrystals occupy random positions in the superlattice provided that they are the same size as the host's nanocrystals (Fig 1a). The chance of each lattice position being occupied by a dopant gold particle can be thought of as a fixed probability, determined by the ratio of the number of gold and semiconductor particles in the suspension from which the superlattice was made. This random occupation results in a uniform distribution, on average, through the superlattice (similar to the atomic dopants in a silicon crystal), and allows the concentration

of the dopant to be gradually increased without causing changes to the structure of the host lattice.

Obtaining such a random lattice occupation is far from a trivial achievement, because two other types of structure could have formed. The first alternative could have occurred through phase segregation: the formation of separate arrays of host and dopant nanocrystals (Fig. 1b). This can occur if the attraction between gold nanocrystals is stronger than that between the host nanocrystals, resulting in crystallization of the former before that of the latter. Cargnello et al. observed that phase segregation does indeed occur if the two types of nanocrystal differ in size. The second alternative is the formation of a binary superlattice — a well-defined crystal structure consisting of two sublattices, in which the two types of nanocrystal occupy specific positions[4,5] (Fig. 1c). In fact, the ordered arrangement of dopants in such binary superlattices can be thought of as being the structural opposite of the random dopant arrangements in the solid solutions reported by Cargnello and co-workers.

Perhaps the best way to understand the observed randomness is to start from a model that considers the nanocrystals as spherical particles that do not attract each other in the suspension. This is a realistic assumption, because the particles are capped with organic molecules that are similar to the hexane molecules used in the starting nanocrystal suspension — which
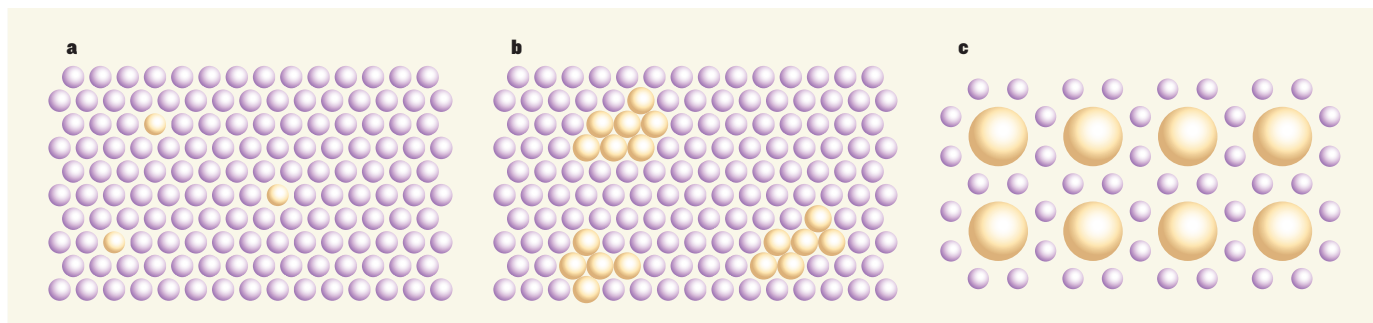

Figure 1 | Alternative outcomes from superlattice doping. Superlattices are arrays of identical nanometre-scale particles. When a few of the particles are replaced by foreign 'dopant' particles, different structural arrangements can occur. **a**, Cargnello et al.[1] report substitutional doping of a cadmium selenide (CdSe) superlattice by gold nanocrystals — random replacement of host particles by dopants — when the gold nanocrystals are identical in size to the CdSe particles. **b**, The authors also observed phase segregation (clustering of the different particle types) when the gold dopants were a different size from the CdSe particles. **c**, Different-sized particles can also form binary superlattices — well-defined crystal structures consisting of two sublattices, in which the two types of nanocrystal occupy specific positions. Binary superlattices did not form in the authors' system.

means that the solvent screens the interactions between the capping molecules of two adjacent nanocrystals, leaving only weak attractions between the particles' cores[6]. If these core–core attractions are either very small or similar in magnitude for the host and the gold nanocrystals, then random mixing without segregation might be expected, provided that both types of nanocrystal are the same size. However, the authors note that a subtle competition between random doping and segregation occurs, depending on the length of the capping molecules, for instance. This suggests that the model described above is too simple.

One of the forces driving the extensive research on nanocrystal superlattices is the prospect of designing materials that have currently unavailable optical, electronic or magnetic properties arising from quantum mechanical[7–9] or dipolar interactions[10] between their building blocks. Cargnello *et al*. find that that the electronic conductivity of the PbSe material increases by a factor of 10^6 when the concentration of dopant gold nanocrystals exceeds 16.5% of the total number of nanocrystals. Their work therefore adds to the growing knowledge in the field by demonstrating for the first time that substitutional doping can indeed enable the manipulation of such properties. ■

Daniel Vanmaekelbergh *is in the Department of Chemistry, Debye Institute for Nanomaterials Science, University of Utrecht, Utrecht 3584 CC, the Netherlands. e-mail: d.vanmaekelbergh@uu.nl*

1. Cargnello, M. *et al. Nature* **524,** 450–453 (2015).
2. Murray, C. B., Kagan, C. R. & Bawendi, M. G. *Science* **270,** 1335–1338 (1995).
3. Vanmaekelbergh, D. *Nano Today* **6,** 419–437 (2011).
4. Shevchenko, E. V., Talapin, D. V., Kotov, N. A., O'Brien, S. & Murray, C. B. *Nature* **439,** 55–59 (2006).
5. Kiely, C. J., Fink, J., Brust, M., Bethell, D. & Schiffrin, D. J. *Nature* **396,** 444–446 (1998).
6. Evers, W. H. *et al. Nano Lett.* **10,** 4235–4241 (2010).
7. Schliehe, C. *et al. Science* **329,** 550–553 (2010).
8. Boneschanscher, M. P. *et al. Science* **344,** 1377–1380 (2014).
9. Zolotavin, P. & Guyot-Sionnest, P. *ACS Nano* **6,** 8094–8104 (2012).
10. Talapin, D. V. *MRS Bull.* **37,** 63–71 (2012).

Surviving import failure

Two studies reveal that dysfunction in organelles called mitochondria causes the toxic accumulation of mitochondrial proteins in the cell's cytosolic fluid, and identify ways in which damage is mitigated. SEE LETTERS P.481 & P.485

COLE M. HAYNES

In an endosymbiotic event that occurred more than one billion years ago, a bacterium was engulfed by a cell, and eventually became an organelle — the mitochondrion. Over time, most of the roughly 1,000 genes that encode mitochondrial proteins were transferred from mitochondria to the nucleus, and are now translated into proteins in the intracellular fluid known as the cytosol. A crucial import mechanism then ensures that these proteins end up in the appropriate locations within mitochondria. Now, two complementary studies[1,2] in this issue provide insight into the consequences of inefficient import of mitochondrial proteins — their accumulation in the cytosol — and demonstrate that the cell undergoes several adaptive responses to mitigate the toxicity caused by such accumulation.

Mitochondria not only act as signalling hubs, but are also responsible for generating most of the cell's energy. Defects in mitochondrial function often arise with ageing, or in diseases associated with neuromuscular degeneration, including Parkinson's disease and amyotrophic lateral sclerosis. In these settings, mitochondrial dysfunction is thought to contribute to cell dysfunction and ultimately death, either by causing abnormal energy production or by initiating a cell-death program known as apoptosis. But could cell death that is related to mitochondrial dysfunction instead arise owing to an unknown or unanticipated effect on other essential cellular compartments or activities?

In the first study, Wang and Chen[1] (page 481) used an unbiased screening approach to identify 40 genes that prevent cell death when overexpressed in cells harbouring damaged mitochondria. None of the proteins encoded by these genes are mitochondrial. Instead, almost all reside in the cytosol, suggesting that mitochondrial dysfunction may alter essential cytosolic functions — a previously unknown effect. Indeed, some of the identified proteins are known to decrease the rate of cytosolic protein synthesis, or to promote protein degradation in the cytosol.

Wrobel and colleagues[2] (page 485) took an alternative approach in the second study, analysing all the RNA transcripts and proteins that are altered in cells in which mitochondrial import is impaired. Strikingly, expression and production of many of the genes and proteins required for protein synthesis were reduced, as was overall protein synthesis. Furthermore, the activity of the proteasome (a large complex that degrades proteins in the cytosol) was increased, as were levels of proteasome assembly factors and chaperone proteins[3,4].

Both studies demonstrate that mitochondrial precursor proteins accumulate in the cytosol when mitochondrial function is perturbed. Interestingly, the proteins are degraded relatively quickly in the cytosol, compared with when they are imported into mitochondria. Combined with the studies' findings that proteasome activity increases when mitochondria are damaged and has a protective role in this setting, these data suggest that the accumulation of mitochondrial proteins in the cytosol
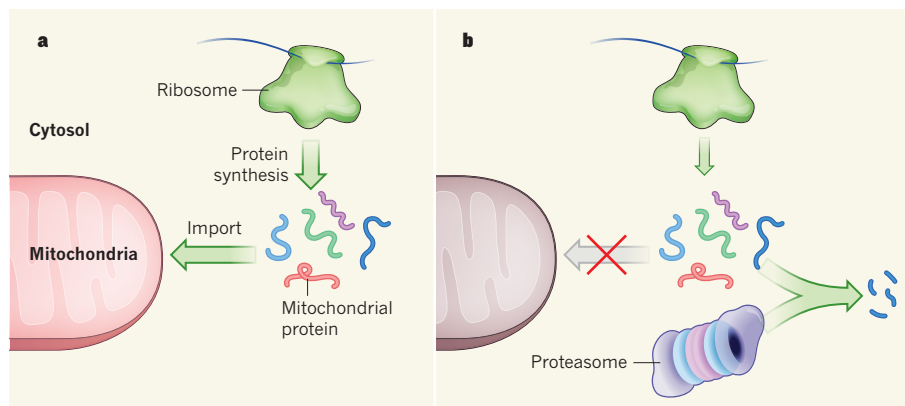

Figure 1 | Effects of decreased mitochondrial-protein import. a, Around 1,000 of the proteins that are synthesized by ribosomes in the cell's cytosol are subsequently imported into organelles called mitochondria. **b,** Mitochondrial dysfunction can impair normal import, leading to toxicity and, if left unchecked, cell death. Two studies[1,2] report that cells limit the accumulation of toxic mislocalized mitochondrial proteins in two ways: by reducing protein synthesis and by increasing the activity of proteasomes, structures that degrade the mislocalized proteins.

can cause death when mitochondria become dysfunctional.

Most mislocalized proteins will retain their signal sequence — a short peptide that is typically removed after it has helped to guide a protein to its proper location in a mitochondrion — and so might be unable to assemble correctly and adopt a functional conformation. And those proteins that are normally integrated into the mitochondrial membranes are probably prone to forming toxic clumps called aggregates in the aqueous environment of the cytosol. The accumulation of such proteins has the capacity to disrupt or overwhelm essential cytosolic activities that are required for general protein synthesis, folding and assembly. These two studies show how cytosolic adaptations reduce the accumulation of mislocalized proteins, allowing cells to better cope with the consequences of mitochondrial dysfunction (Fig. 1).

What are the underlying mechanisms by which cytosolic protein degradation is increased and synthesis is decreased when mitochondrial proteins accumulate in the cytosol? Wang and Chen provide part of the answer, showing that the mitochondrial proteins somehow stabilize cellular protein components that are known[3,5] to reduce protein synthesis. Normally, these components are rapidly degraded by proteasomes. But, in the presence of mislocalized mitochondrial proteins, the components avoid degradation, accumulate and reduce protein synthesis.

Both studies reported an increase in proteasome assembly factors, which seems to be independent of transcription. But the mechanism by which levels of these proteins are increased remains unclear. Furthermore, although Wrobel and colleagues reported that changes in the transcription of some genes are required for cells to survive when import is impaired, the way in which this change is regulated remains to be discovered. However, it does not seem to require the signalling mechanisms known[6] to be associated with the accumulation of misfolded or aggregated cytosolic proteins, such as the heat-shock response.

Perhaps most importantly, how are the mislocalized mitochondrial proteins identified or detected in the cytosol and directed to the proteasome? Wrobel and colleagues demonstrate that mislocalized proteins are marked with the protein ubiquitin, which tags them for degradation by proteasomes. This indicates that a currently unknown ubiquitin ligase enzyme is involved in their degradation. But the features that indicate that a protein is mislocalized are unclear, because neither the addition of ubiquitin tags nor proteasomal degradation require the mitochondrial signal sequence.

Going forward, it will be interesting to understand how the newly discovered stress response interacts with the two other responses activated by mitochondrial dysfunction and impaired mitochondrial import: a transcriptional response known as the mitochondrial unfolded protein response and the mitophagy pathway. The mitochondrial unfolded protein response is activated to promote survival and mitochondrial repair during mitochondrial dysfunction[7]. By contrast, the mitophagy pathway uses impaired mitochondrial import to recognize the most severely damaged organelles, and then degrades them to improve cellular fitness[8].

Finally, at what point does the toxicity of mislocalized mitochondrial proteins engage apoptotic cell-death pathways? As the authors of both papers suggest, an understanding of this interaction could help to pave the way for the development of treatments for mitochondrial diseases, which until now were thought to arise predominantly from defects in energy production. Perhaps therapeutic strategies to treat mitochondrial diseases should focus on remedying the cytosolic defects caused by mitochondrial-protein accumulation. In support of this suggestion, mice that are treated with a compound that reduces protein synthesis are protected against mitochondrial disease[9], providing a cause for optimism about future treatments. ∎

Cole M. Haynes *is at the Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.*
e-mail: haynesc@mskcc.org

1. Wang, X. & Chen, X. J. *Nature* **524,** 481–484 (2015).
2. Wrobel, L. *et al. Nature* **524,** 485–488 (2015).
3. Matsuo, Y. *et al. Nature* **505,** 112–116 (2014).
4. Le Tallec, B. *et al. Mol. Cell* **27,** 660–674 (2007).
5. Sammons, M. A., Samir, P. & Link, A. J. *Biochem. Biophys. Res. Commun.* **406,** 13–19 (2011).
6. Akerfelt, M., Morimoto, R. I. & Sistonen, L. *Nature Rev. Mol. Cell Biol.* **11,** 545–555 (2010).
7. Nargund, A. M., Pellegrino, M. W., Fiorese, C. J., Baker, B. M. & Haynes, C. M. *Science* **337,** 587–590 (2012).
8. Narendra, D. P. *et al. PLoS Biol.* **8,** e1000298 (2010).
9. Johnson, S. C. *et al. Science* **342,** 1524–1528 (2013).

This article was published online on 5 August 2015.

STRUCTURAL BIOLOGY

Lipid gymnastics

Crystal structures of the bacterial protein PglK uncover structural features that suggest how the protein 'flips' lipid–bound oligosaccharide molecules from one side of the cell membrane to the other. See Article p.433

ALICE VERCHÈRE & ANANT K. MENON

Cell-surface and secreted proteins are almost always decorated with sugar chains called oligosaccharides. In some types of this glycosylation process the oligosaccharides are first assembled on a lipid carrier molecule, before being transferred en bloc to proteins. A striking feature of lipid-linked oligosaccharide (LLO) biosynthesis is that it is initiated on the cytoplasmic side of intracellular membranes but the attachment of the oligosaccharides to proteins occurs on the opposite side. Thus, the LLO must be transported — flipped — across the membrane. On page 433 of this issue, Perez *et al.*[1] reveal how a bacterial 'flippase' protein might accomplish this extraordinary feat.

Protein glycosylation occurs in all domains of life[2]. *N*-Glycoproteins, in which an oligosaccharide is linked to an asparagine amino acid, are synthesized in a cellular organelle called the endoplasmic reticulum (ER) in eukaryotic organisms (those with nucleated cells), and in the periplasm of bacteria — the space between the cell's cytoplasmic and outer membranes. Although distinct pathways are used by both classes of organism, they share two defining features. First, an LLO intermediate is assembled on the cytoplasmic face of the ER or periplasmic membrane from sugar units and an unusual lipid: a very long, highly flexible[3] polyisoprene chain. Second, the LLO must be flipped across the membrane.

Because LLO flipping is an energetically costly manoeuvre, it has always been assumed that the task falls to specific membrane-spanning proteins called flippases. In eukaryotes, transfer of LLO across the ER membrane occurs without the use of ATP[4,5] (the 'currency' molecule of cellular energy), but no transporter protein has yet been identified[6]. By contrast, in bacteria, the membrane protein PglK has been proposed[7] to use energy provided by ATP hydrolysis to move LLOs from the cytoplasmic to the periplasmic side of the membrane.

Perez *et al.* provide definitive evidence that PglK is an LLO flippase. By incorporating purified PglK into lipid vesicles containing a synthesized LLO intermediate (tLLO), they demonstrate ATP-dependent, PglK-dependent flipping of tLLO from the outer leaflet to the interior of the vesicles — corresponding to from the cytoplasm to the periplasm. They also show that flipping does not occur if PglK's ability to hydrolyse ATP is disabled. It has long been known that the ATP-hydrolysing activity of proteins of the ABC transporter family is stimulated by their normal substrate, and PglK
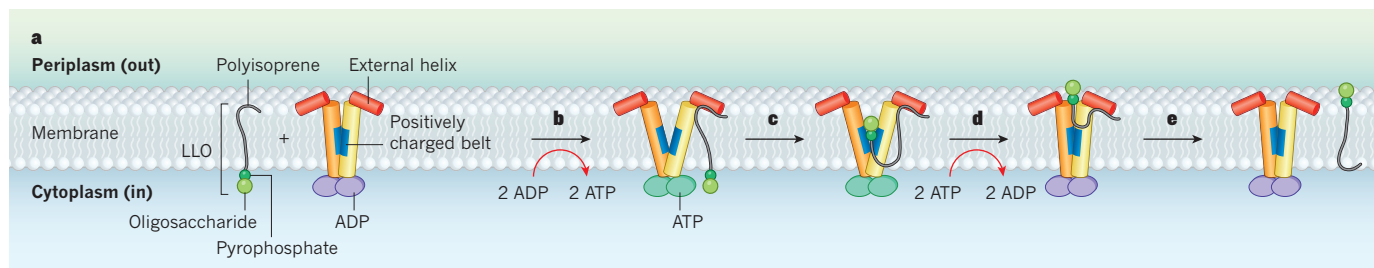
Figure 1 | Proposed PglK flippase mechanism. **a**, PglK is a homodimeric protein that spans the membrane between the cytoplasmic and periplasmic compartments in bacterial cells. The protein is responsible for 'flipping' a lipid-linked oligosaccharide (LLO) molecule, which comprises an oligosaccharide that is linked by a negatively charged pyrophosphate to a lipid (polyisoprene) tail. Perez et al.[1] present crystal structures of PglK showing that each monomer has a binding site for an ATP molecule (the cellular energy currency) or its hydrolysed form ADP, a belt of positively charged amino acids and an external helix on the outward (periplasmic) side. The flipping cycle begins with the dimer in an outward-occluded conformation and the oligosaccharide head of LLO on the inward (cytoplasmic) side. **b**, The polyisoprene tail of LLO binds a region of PglK that includes the external helix, promoting the exchange of bound ADP for ATP and generating the outward-open state of the protein. **c**, The pyrophosphate–oligosaccharide head then enters the V-shaped cavity, attracted by its positively charged lining. **d**, On ATP hydrolysis, PglK returns to an outward-occluded conformation that squeezes out the LLO head region to the outward side of the membrane. **e**, The LLO tail then disengages from PglK to extend to the cytoplasmic side of the membrane, so that the LLO is in the opposite orientation from that at the beginning of the cycle.

is no exception. Perez et al. show that LLOs stimulate PglK's ATPase activity, reinforcing the idea that these molecules are the protein's natural substrates.

The authors also present three crystal structures of PglK, which is a homodimeric protein. Two of these correspond to the protein with no ATP or ADP (the hydrolysed form of ATP) bound. These structures show a V-shaped conformation with the mouth of the V opening into the cytoplasm. Similar structures have been reported for other ABC transporters, and this cavity has been suggested[8] to be a substrate-binding pocket. However, because there is plenty of ATP and ADP in the cytoplasm, it is unlikely that these conformations would be prominent under physiological conditions. Indeed, the authors show that mutating PglK to reduce the cavity volume had no effect on LLO flipping or the protein's ATPase activity, arguing against a role for this cavity in substrate binding. The authors' third structure, with ADP bound, shows an outward-facing (periplasm-facing) V, intermediate in shape between the completely occluded and completely open structures determined for other ABC transporters.

The structures also reveal two unusual features. The first is a short α-helix at the periplasmic side of each PglK monomer, situated parallel to the plane of the membrane. Together with specific transmembrane segments, these external helices form hydrophobic grooves near the membrane boundary. The authors show that mutating these helices severely reduces rates of LLO flipping and prevents the stimulation of the protein's ATPase activity by LLO. The second feature is a belt of positively charged amino acids lining the interior of the outward-facing cavity. This region could interact with the negatively charged pyrophosphate linker in the LLO headgroup. A PglK mutant without this belt had no flippase activity and its residual ATPase activity could also no longer be stimulated by LLO.

The authors use these structural features to propose a mechanism for LLO flipping (Fig. 1). The polyisoprene tail of LLO inserts into the hydrophobic groove created by the external helices, and the molecule's oligosaccharide headgroup enters the outward-facing cavity of the ATP-bound protein to interact with the belt of positive charges. As ATP is hydrolysed to ADP, the protein switches to the occluded structure, causing the headgroup to be squeezed outwards. Closure of the outward-facing V conformation prevents re-entry of the molecule.

This model suggests that the inward-facing conformation of PglK plays no part in transport, which is a departure from the mechanism proposed[9] for the bacterial flippase MsbA, in which the entire lipid enters the inward-facing cavity in the first step of transport. It also suggests that the pyrophosphate linker, rather than the oligosaccharide, is the key structural feature of the LLO headgroup, which is consistent with PglK's lack of selectivity for the LLOs that it can flip. Furthermore, the mechanism relies on the flexibility of the LLO's polyisoprene tail, which spans the membrane at the start of the transport sequence yet loops fully around so that both head and tail are on the same side of the membrane at the end.

As with any innovative proposal, Perez and colleagues' model is thought-provoking. Although the authors' data on PglK variants with mutated external helices suggest that this region binds the polyisoprene tail, there is no direct evidence of binding, and it is not clear how the tail comes unstuck from these grooves once flipping is complete. Furthermore, the identification of one external helix per PglK monomer indicates that the homodimeric protein might bind two LLOs simultaneously. This possibility is supported by hints of cooperativity in the stimulation of the protein's ATPase activity by LLOs, but also remains to be verified. Another major question is how does the pyrophosphoryl oligosaccharide headgroup enter the outward-facing cavity? This must occur through a gap, permanent or transient, near the cytoplasmic side of PglK, but this is not evident in the available structure. Could it be that transient, inward-facing conformations allow initial entry of the headgroup? Even if this is the case, the outward-facing structures must still contain an aperture to allow penetration of the polyisoprene lipid. These issues may only be resolved by obtaining LLO-bound structures of PglK.

Although many details of the proposed PglK flippase mechanism remain to be evaluated, it is interesting to ponder its generality. The requirement for a long, flexible lipid component may make the PglK mechanism specific to LLO transporters. For example, MsbA probably functions through a different mechanism[9], because the molecule this protein flips, Lipid A, has only the usual type of fatty-acyl chains[10]. It is also worth considering the LLO-transporting scramblase proteins that are needed for N-glycosylation in the eukaryotic ER and for synthesizing the bacterial cell wall[5]. In these cases, LLO transport is ATP-independent and bidirectional. By analogy with the proposed role of the external helices in PglK, perhaps a surface feature on the transmembrane portion of the scramblase provides an attachment point for the polyisoprene tail. The polyisoprene-binding motifs found in many proteins involved in glycosylation pathways may constitute precisely such an attachment site[11]. ■

Alice Verchère and **Anant K. Menon** are in the Department of Biochemistry, Weill Cornell Medical College, New York, New York 10065, USA.
e-mail: akm2003@med.cornell.edu

1. Perez, C. et al. Nature **524,** 433–438 (2015).
2. Larkin, A. & Imperiali, B. Biochemistry **50,** 4411–4426 (2011).
3. Kern, N. R. et al. Biophys. J. **107,** 1885–1895 (2014).
4. Schenk, B., Fernandez, F. & Waechter, C. J. Glycobiology **11,** 61R–70R (2001).
5. Sanyal, S. & Menon, A. K. ACS Chem. Biol. **4,** 895–909 (2009).
6. Frank, C. G., Sanyal, S., Rush, J. S., Waechter, C. J. & Menon, A. K. Nature **454,** E3–E4 (2008).
7. Alaimo, C. et al. EMBO J. **25,** 967–976 (2006).

8. ter Beek, J., Guskov, A., & Slotboom, D. J. *J. Gen. Physiol.* **143,** 419–435 (2014).
9. Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. *Proc. Natl Acad. Sci. USA* **104,** 19005–19010 (2007).
10. Doerrler, W. T., Gibbons, H. S. & Raetz, C. R.

J. Biol. Chem. **279,** 45102–45109 (2004).
11. Zhou, G.-P. & Troy, F. A. 2nd. *Glycobiology* **13,** 51–71 (2003).

PARTICLE PHYSICS

Positrons ride the wave

Experiments reveal that positrons — the antimatter equivalents of electrons — can be rapidly accelerated using a plasma wave. The findings pave the way to high-energy electron–positron particle colliders. SEE LETTER P.442

PHILIPPE PIOT

Particle accelerators and colliders have been the backbone of research into elementary particle physics for almost a century. The next generation of colliders must be able to generate collision energies for lepton particles, such as electrons and positrons, in the teraelectronvolt range (1 TeV is 10^{12} electronvolts). This will complement discoveries being made by the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. On page 442 of this issue, Corde *et al.*[1] report a system for accelerating positrons — the antimatter counterparts of electrons — that might enable such instruments to be realized.

An experimental configuration of choice for colliders consists of smashing electrons (e^-) into positrons (e^+). The two particle beams are accelerated in two linear accelerators pointing towards each other, to collide head-on. The designs of e^+–e^- linear colliders have mostly been based on conventional accelerator technologies that use devices known as radio-frequency resonators. But these have allowed accelerating gradients of no more than about 100 MeV per metre. The TeV-energy colliders are therefore expected to be large and costly — for example, the proposed 0.5-TeV International Linear Collider is expected to be about 30 km long[2].

Alternative acceleration methods have been sought since the early 1980s (ref. 3). One class of such methods is plasma wakefield beam-driven acceleration (PWFA), in which the accelerating medium is an electrically neutral plasma composed of positively charged ions and highly mobile free electrons. As a group of electrons called the drive bunch travels through the plasma, it repels the plasma electrons. This leads to spatial modulation of the plasma's electronic density, which in turn induces a strong electric field commonly called a wakefield. When another group of electrons — the witness bunch — travels at an appropriate distance behind the drive bunch, it interacts with the wakefield. This interaction causes the witness bunch to accelerate rapidly.
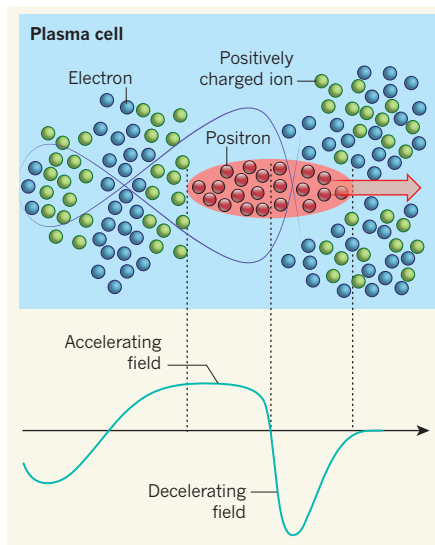

Figure 1 | Mechanism for positron acceleration. Corde *et al.*[1] report that positrons can be accelerated in a lithium plasma composed of positive ions and free electrons; only some of the ions and electrons are shown. As the positron bunch moves from left to right, the electrons move towards it because of charge attraction; the blue line indicates a sample electron trajectory. This movement organizes electrons and ions into 'layers' behind the bunch, generating an electric field called a wakefield (depicted as a cyan trace on the lower plot). The positrons at the head of the bunch decelerate in response to the negative region of the wakefield, but positrons in the bunch's tail accelerate in an approximately uniform, positive region of the field.

The PWFA method proposed[4] in the mid-1980s for electron acceleration was initially developed in the linear regime, in which the electron beam has a lower particle density than the plasma. It was subsequently realized that this approach could also operate in a nonlinear regime (in which the electron beam has a higher particle density than the plasma), giving rise to sharp transitions in the plasma-density modulations that could support higher accelerating fields[5]. Using the latter mechanism, often referred to as a blowout regime, accelerating gradients of about 52 GeV m^{-1}

have been attained for electron bunches[6].

Application of the PWFA technique to accelerate positrons over a short distance is a prerequisite for a plasma-based e^+–e^- linear collider. Previous work[7] in this area demonstrated energy modulation of a positron bunch with a peak accelerating gradient of the order of 50 MeV m^{-1}. The possibility of accelerating positrons in the nonlinear regime with high acceleration has long been contemplated, but not achieved. The main hurdles arise because the dynamics of plasma electrons under the influence of a positron drive bunch are different from the dynamics induced by an electron drive bunch, and because electron bunches are not suitable for driving plasma waves that can accelerate positron bunches.

Corde *et al.* now demonstrate positron acceleration with accelerating gradients of about 5 GeV m^{-1}, which is about 100 times higher than the gradients previously reported for positrons[7]. In brief, the authors used a conventional accelerator to produce a positron bunch, and sent it through a metre-long cell of lithium plasma.

The plasma-density wave formed using positrons is different from that produced using electrons because the positron bunch attracts plasma electrons inward. As electrons flow through the bunch, the positrons gain or lose energy depending on their axial locations within the bunch (Fig. 1). At a particular plasma density, the wakefield can be tailored to switch polarity within the positron bunch, thereby decelerating the head of the bunch while accelerating positrons in the tail, effectively transferring energy from the head to the tail.

Corde *et al.* observe that the large positron population (about 1 billion positrons) experiencing the accelerating field effectively 'loads' the wakefield and affects its shape, leading to an approximately uniform energy gain for the accelerated positrons. The experiment therefore demonstrates that, by using appropriate operating parameters, only one positron bunch is needed for acceleration: part of its trailing population is 'trapped' and accelerated quasi-uniformly to higher energies, and so splits from the initial bunch. The authors observe that the newly formed and the accelerated positron bunches are quasi-monoenergetic — the positrons in the accelerated bunch have almost the same energy as each other. This is essential if the bunches are to be further transported and manipulated along an accelerator without substantial degradation of quality, for example before injection into a subsequent stage of a plasma wakefield accelerator.

Several challenges must be addressed to prove the overall suitability of a PWFA-based e^+–e^- linear collider. First, several of Corde and colleagues' PWFA modules must be concatenated to reach the desired final energy. The quasi-monoenergetic bunch formed in the first module would have to be used as a witness bunch by subsequent modules, and so the

related synchronization issue — how to delay the witness bunch behind the drive bunch with sub-picosecond accuracy so that it interacts properly with the wakefield — would have to be investigated. It will also be important to understand and mitigate possible beam-quality degradation during the acceleration process as particles in the accelerated bunches scatter against the plasma electrons (and ions). Several

groups are already vigorously investigating these issues; Corde and colleagues' results provide further impetus to these studies. ∎

Philippe Piot *is in the Department of Physics, Northern Illinois University, DeKalb, Illinois 60115, USA, and at the Fermi National Accelerator Laboratory, Batavia, Illinois.*
e-mail: ppiot@niu.edu

1. Corde, S. *et al. Nature* **524,** 442–445 (2015).
2. ILC Technical Design Report. www.linearcollider. org/ILC/Publications/Technical-Design-Report
3. Joshi, C. & Katsouleas, T. *Physics Today* **56**(6), 47–53 (2003).
4. Chen, P., Dawson, J. M., Huff, R. W. & Katsouleas, T. *Phys. Rev. Lett.* **54,** 693–696 (1985).
5. Rosenzweig, J. B., Breizman, B., Katsouleas, T. & Su, J. J. *Phys. Rev. A* **44,** R6189–R6192 (1991).
6. Blumenfeld, I. *et al. Nature* **445,** 741–744 (2007).
7. Blue, B. E. *et al. Phys. Rev. Lett.* **90,** 214801 (2003).

EVOLUTION

Gene transfer in complex cells

A comparative genomic study shows that, during evolution, nucleus-containing cells acquired DNA from bacteria primarily by endosymbiosis — the uptake and integration of one cell by another. SEE ARTICLE P.427

JOHN M. ARCHIBALD

Once controversial, the idea that genetic material can be transferred laterally between organisms is now known to be a key factor in the evolution of the prokaryotes (bacteria and archaea), whose DNA is not enclosed in a nucleus[1]. However, it is unclear to what extent such transfer affects eukaryotic (nucleus-bearing) cells, which are typically thought to transmit their genes vertically from parent to offspring. In this issue, Ku *et al.*[2] (page 427) assess the contribution of lateral gene transfer (LGT) to the eukaryotic nuclear genome, and conclude that, although prokaryote-to-eukaryote LGT has happened, it has not extensively affected the eukaryotic cell over long evolutionary timescales.

The nuclear genome is an ever-evolving mosaic of DNA acquired from different sources (Fig. 1). For instance, mitochondria and chloroplasts are eukaryotic organelles derived from prokaryotic cells that were assimilated by another cell through a process called endosymbiosis. Although these organelles usually retain some DNA, most of their original genome has moved to the host nucleus through endosymbiotic gene transfer (EGT)[3]. In the case of chloroplasts, subsequent eukaryote–eukaryote endosymbioses have spread the organelle and its associated genes laterally across the evolutionary tree[4]. But, endosymbiosis aside, how web-like is eukaryotic genome evolution?

In multicellular eukaryotes, the separation of the sex cells from the rest of the organism is often assumed to be a strong barrier to the stable acquisition of foreign DNA. However, the strength

of this assumption has been questioned[5], and there are compelling examples of recent, lineage-specific LGT in both animals[6] and plants[7]. Viruses can serve as gene-transfer vectors[8] and might facilitate LGT in eukaryotes as they do in prokaryotes[1]. Single-celled eukaryotes might acquire DNA simply by ingesting and digesting prey[9]. Most eukaryotic LGT has been inferred solely on the basis of phylogenetic incongruence — genes whose evolutionary histories are at odds with known or predicted organismal relationships. But
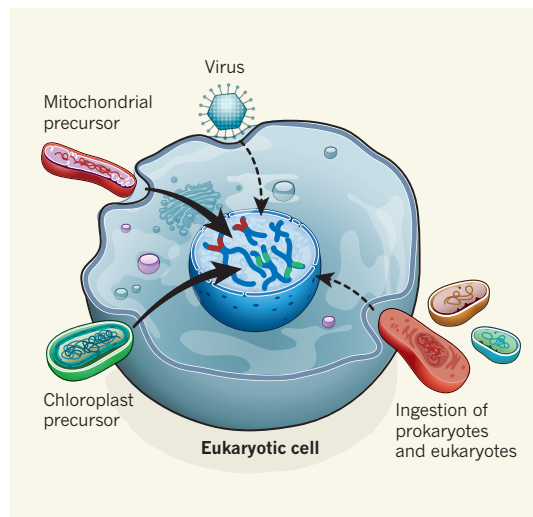

Figure 1 | Acquiring foreign DNA. The nuclear genome of eukaryotic (nucleus-containing) cells is a mosaic of genes from various sources, including prokaryotes (bacteria and archaea). Foreign DNA may enter eukaryotes when they ingest other cells, or through viruses, which can act as gene-transfer vectors. Ku *et al.*[2] report that such lateral gene transfer has had a minimal impact on the nuclear genome (indicated by dashed arrows), but that the origin of organelles called mitochondria and chloroplasts from bacterial endosymbionts provided a rich source of foreign DNA.

phylogenetic trees are prone to artefacts and open to interpretation. As such, the true impact of LGT in eukaryotes is debated.

Ku *et al.* studied thousands of genomes from across the breadth of cellular life. Their analytical framework is based on the following premise: if prokaryote-to-eukaryote LGT has played a meaningful part in eukaryotic evolution, then its long-term impact should be detectable and cumulative. Nuclear genomes should accumulate prokaryotic genes and become increasingly different from one another over time. By contrast, the accumulation of prokaryotic genes due to EGT will be episodic.

The authors compared just over 950,000 protein sequences inferred from the genomes of 55 different eukaryotes to one another and to more than 6 million prokaryotic sequences. The analysis identified 2,585 sequence families, each comprising related (homologous) proteins encoded by two or more eukaryotic genomes and five or more prokaryotic genomes: these families were designated as eukaryote–prokaryote clusters. The remaining 26,117 families were labelled eukaryote-specific clusters.

An overview of the distribution of eukaryote–prokaryote clusters reveals an expectedly large evolutionary footprint associated with the endosymbiotic origin of chloroplasts. Hundreds of nuclear genes are absent in species lacking chloroplasts but are shared by various groups of photosynthetic eukaryotes (and are abundant in cyanobacteria, from which chloroplasts evolved[4]). The endosymbiotic footprint of mitochondrial evolution is also apparent, albeit less distinct.

Phylogenetic trees inferred from each eukaryote–prokaryote cluster paint a remarkable picture of the history of prokaryotic genes in eukaryotes. For 74.8% of these clusters, the proteins from different eukaryotes are on adjacent branches (they are monophyletic), and eukaryote monophyly cannot be ruled out for an additional 12.7% of protein families. This pattern would not be expected if eukaryotes were steadily acquiring genes from different prokaryotic groups.

Instances of eukaryote monophyly could, nevertheless, be the result of a single prokaryote-to-eukaryote LGT

followed by multiple LGT events between eukaryotes. Indeed, eukaryotic homologues in the eukaryote–prokaryote clusters are often patchily distributed: some are present in only two of the six recognized high-level groups of eukaryotes, and many more are limited to three or four such groups. Although patchy gene distributions are often interpreted as evidence for LGT[1,5,10], Ku *et al.* conclude that the trees in which eukaryotic sequences are monophyletic are not the product of eukaryote-to-eukaryote LGT. Why? Statistical tests showed that the trees for eukaryote–prokaryote clusters are generally compatible with those inferred from the eukaryote-specific clusters, which are considered to be the product of vertical evolution.

The remaining 12.5% of trees for eukaryote–prokaryote clusters show the eukaryotic homologues branching apart from one another, consistent with prokaryote-to-eukaryote LGT. Ku *et al.* provide alternative explanations for such patterns, including sequencing contaminations and errors inherent in phylogenetic reconstructions. However, the authors identified several blocks of genes whose lineage-specific distributions are so striking that LGT is the only reasonable explanation. It will be interesting to see how these LGTs stack up against those identified by others[11].

Ku *et al.* conclude that gene evolution in eukaryotes is "resoundingly vertical" and that the punctate distribution of prokaryotic genes across eukaryotes is primarily the result of differential gene loss during evolution. Apart from the gene acquisitions associated with mitochondrial and chloroplast EGT, eukaryotes seem to sample prokaryotic gene diversity at a much lower level than do bacteria and archaea. Given evidence for at least some prokaryote-to-eukaryote LGT, the authors suggest that perhaps genes transferred by LGT are retained in the genome for only a short time, or that lineages that engage in LGT tend not to be successful in the long run. This latter idea is at odds with the prevailing view that LGTs are of benefit to the recipient organism[1,10].

At present, there are two types of eukaryotic gene whose histories are uncontroversial: ubiquitous genes that were probably present in the common ancestor of all eukaryotes; and lineage-specific genes with strong signatures of recent LGT. Between these two extremes lies a continuum of genes whose phylogenies and presence–absence patterns are exceedingly complex. The extent to which Ku and colleagues' analyses have fully captured the evolutionary forces that shaped the nuclear genome is unclear. Nonetheless, they have shown that EGT has been the dominant mode of gene acquisition in eukaryotes, that gene patchiness is the norm and that gene loss needs to be taken seriously. We now have the opportunity to compare

notes on the strengths and weaknesses of the various approaches currently being used to distinguish lateral from vertical inheritance in eukaryotes. Consider the issue open for discussion. ∎

John M. Archibald *is in the Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada, and at the Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Toronto.*
e-mail: john.archibald@dal.ca

1. Soucy, S. M., Huang, J. & Gogarten, J. P. *Nature Rev. Genet.* **16,** 472–482 (2015).
2. Ku, C. *et al. Nature* **524,** 427–432 (2015).
3. Martin, W., Brinkmann, H., Savonna, C. & Cerff, R. *Proc. Natl Acad. Sci. USA* **90,** 8692–8696 (1993).
4. Gould, S. B., Waller, R. F. & McFadden, G. I. *Annu. Rev. Plant Biol.* **59,** 491–517 (2008).
5. Huang, J. *BioEssays* **35,** 868–875 (2013).
6. Boto, L. *Proc. R. Soc. B* **281,** 20132450 (2014).
7. Rice, D. W. *et al. Science* **342,** 1468–1473 (2013).
8. Mi, S. *et al. Nature* **403,** 785–789 (2000).
9. Doolittle, W. F. *Trends Genet.* **14,** 307–311 (1998).
10. Keeling, P. J. *Curr. Opin. Genet. Dev.* **19,** 613–619 (2009).
11. Alsmark, C. *et al. Genome Biol.* **14,** R19 (2013).

PHOTONICS

A stable narrow–band X–ray laser

An atomic laser operating at the shortest wavelength yet achieved has been created by bombarding a copper foil with two X-ray pulses tuned to slightly different energies. The results may lead to ultrastable X–ray lasers. SEE LETTER P.446

LINDA YOUNG

X-rays penetrate matter to image a system's internal 3D structure using contrast arising from spatial variations of elemental, chemical or magnetic properties. Lasers that function at X-ray wavelengths go beyond basic structure determination, because they can deliver bright pulses on ultrashort timescales to probe matter at the atomic level. This means that they can be used to characterize dynamic processes, such as chemical-bond formation, charge transfer and light-induced superconductivity, or to determine the macromolecular structure of a system without damaging it. Such lasers have been the subject of extreme fascination since physicist Theodore Maiman demonstrated[1] the first laser that operated at optical wavelengths in 1960. On page 446 of this issue, Yoneda *et al.*[2] demonstrate an atomic X-ray laser that yields a marked improvement in wavelength stability compared with X-ray free-electron lasers (XFELs), taking a major step towards an ångström-wavelength laser that remains in phase over its pulse duration — that is, which possesses longitudinal coherence.

XFELs[3–5], which use a high-energy electron beam as the laser-generating medium, have revolutionized X-ray science by introducing ultra-fast, ultra-intense X-ray pulses suitable for a vast range of applications. These facilities accelerate electron beams close to the speed of light (at energies in excess of 10^9 electron-volts), through 100-metre-long arrays of magnets that are arranged in a periodic pattern of alternating polarity along the beam path.

These systems create, in a single passage of the electron beam, intense X-ray pulses lasting only a few femtoseconds (1 fs is 10^{-15} seconds) that contain a trillion X-ray photons and achieve peak brightness a billion times greater than radiation produced by conventional synchrotron light sources. XFELs typically operate on the principle of self-amplified spontaneous emission (SASE), whereby the accelerating electrons' incoherent emission of radiation is further amplified by continuous interaction with the electron beam over the length of the magnetic array.

Although SASE results in intense, short-wavelength laser pulses that are coherent in a plane transverse to the direction of propagation, strong longitudinal fluctuations in the time and spectral domains are observed; these can greatly complicate experiments that use these lasers. Yoneda and colleagues' X-ray laser amplifies light that has a well-defined wavelength of 1.54 Å, generated by transitions of electrons from the $2p$ orbital to the $1s$ orbital of copper atoms. This is the shortest-wavelength atomic laser ever demonstrated, surpassing by a factor of 10 an atomic neon laser that has been shown[6] to operate at 14.6 Å.

The authors' laser uses the photoionization-pumping scheme presciently proposed[7] for copper in 1967. In this scheme, the ejection of an inner-shell electron of a copper atom by a suitable light source (the pump) leaves a deficit of electrons in a lower energy level of the resulting copper ion, achieving what is known as population inversion. Electrons drop into the vacated level, spontaneously emitting photons, which are amplified as they
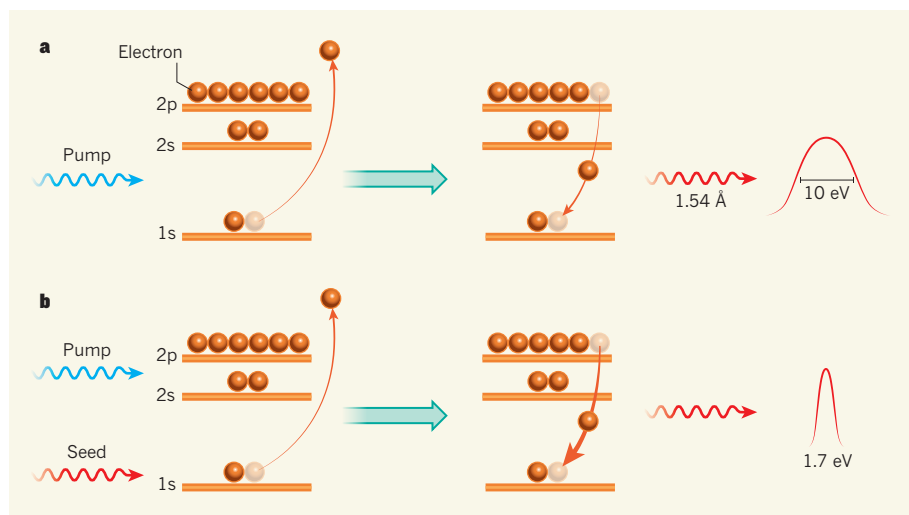
Figure 1 | An atomic laser that uses copper's inner-shell electrons. Yoneda *et al.*[2] report an X-ray laser that operates by focusing intense pulses from an X-ray free-electron laser on copper foil. **a**, They first use a 'pump' pulse that ejects an electron from the copper atoms' 1*s* ground level. This leaves a surplus of electrons in the upper 2*p* energy level of the resulting ions, achieving a state called population inversion. 2*p* electrons then spontaneously drop into the vacated 1*s* level, emitting 1.54-Å photons and starting a process called amplified spontaneous emission (ASE). The resulting laser output has a large spectral linewidth (about 10 eV), which increases with the intensity of the pump. **b**, The authors subsequently used an additional X-ray pulse called the seed, tuned to the lasing wavelength of 1.54 Å. This enhanced (thick orange arrow) the stimulated emission from the upper energy levels. The resulting laser has a much narrower spectral profile (1.7 eV) than the ASE laser and offers increased wavelength stability.

propagate along a path of other identical ions that have inverted electron populations, in a process known as amplified spontaneous emission (ASE). However, Yoneda *et al.* go beyond a simple observation of ASE to demonstrate the 'seeded' operation of the laser that offers improved performance.

The authors first demonstrate amplification of the 1.54-Å transition by focusing brief (7 fs), 9-keV-energy XFEL pulses to a 100-nm-wide spot on a 20-μm-thick copper foil. These are known as pump pulses, and they preferentially eject a 1*s*-orbital electron to ionize the copper atoms in the foil and produce a wave of ions with inverted electron populations that travels collinearly with the generated ASE (Fig. 1a). For a 2-fold increase in the XFEL's power, the authors observe a roughly 15-fold increase in the intensity of the output lasing emission, although no saturation of the laser was observed over the accessible range of pump power — the energy output of the laser never levelled off with increasing pump power.

The authors then demonstrate the seeded operation of the atomic laser (Fig. 1b). For this, they use two XFEL pulses: the pump and another known as the seed, which have energies of 9 and 8 keV, respectively. These are generated by a single electron bunch, ensuring temporal and spatial overlap between the respective pulses. The authors observe that the output of the seeded laser has a narrower linewidth (1.7 eV) than the ASE laser generated using only the pump source, whose linewidth unexpectedly broadens at high pump powers. Thus, the seeded laser output is spectrally more coherent and offers increased

wavelength stability. The seeded laser also represents a substantial improvement compared with the 40-eV-wide input pump pulse.

Why did we have to wait more than 50 years after Maiman's achievement for an atomic X-ray laser operating at such a short wavelength? The main obstacle has been that the power of the pumping source required to create population inversion scales inversely with the fourth power of the output lasing wavelength. This simple consideration necessitates an approximately 10^{15}-fold increase in pump power density to go from a laser operating at 7,000 Å to one at 1.54 Å.

Before XFELs, there were many approaches to achieving such power densities, including the nuclear-bomb-pumped lasers[8] in the US Strategic Defense Initiative programme of the 1980s and optical (or infrared) pumping[9,10]. In the optical-pumping scheme, a 532-nm-wavelength laser was focused on a thin foil of selenium to create a plasma of highly charged ions and electrons[9]. Collisions between electrons and selenium that had lost its outer electrons populated the upper level of the 3*p*-to-3*s* transition of the ions and resulted in a 209-Å laser. By contrast, the photo-ionization-pumping scheme[7] used in the current work directly creates population inversion in a single step by ejecting an inner 1*s* electron, but the method had to wait for the development of XFELs to generate the requisite X-ray power density.

Despite this giant step in X-ray-laser development by Yoneda *et al.*, their atomic copper laser is not quite ready for general use. Their seeded laser did not achieve saturation,

and the absolute energy, angular divergence and duration of the output pulse were not measured. The authors' simulations predict that the conversion efficiency of the laser, which is the ratio of the output power to the input power, is only 2%, corresponding to laser energies of about 1 microjoule. For comparison, the atomic neon laser[6] eventually achieved saturation with about 10% conversion efficiency[11] and yielded energies of about 15 μJ.

Considering slightly longer wavelengths, there are viable alternatives for users seeking the absolute wavelength stability and coherence of atomic lasers. Pulses from tabletop optical lasers can be used as pumps to generate soft X-ray atomic lasers, which operate on the 4*d*-to-4*p* transition of elements with a nickel-like electronic configuration. These lase at 7.3 nm with a power of several microjoules per pulse[12], albeit at a low pulse-repetition rate of 1 hertz. Recently, a tabletop atomic laser with a high average power (1 μJ per pulse; 100 Hz) was demonstrated at 13.9 nm (ref. 13). Tabletop infrared lasers generating 'high-harmonic' emission can provide ultrafast coherent X-ray pulses in a slightly different parameter space (sub-femtosecond duration, broad bandwidth with lower X-ray pulse energies[14]) compared with the optically pumped, atomic X-ray lasers[12,13].

For those seeking improved longitudinal coherence together with narrow-bandwidth and wavelength-tunable X-ray radiation, various seeding schemes have been demonstrated using XFELs, including self-seeding in the high-energy X-ray range[15] and high-gain harmonic generation at the FERMI[16] facility in Trieste, Italy. Yoneda and colleagues' work, together with other improvements to X-ray lasers beyond SASE radiation, ensure that researchers can look forward to a rich playground of very-short-wavelength sources. ■

Linda Young *is in the X-ray Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA.*
e-mail: young@anl.gov

1. Maiman, T. H. *Nature* **187,** 493–494 (1960).
2. Yoneda, H. *et al. Nature* **524,** 446–449 (2015).
3. Ackermann, W. *et al. Nature Photon.* **1,** 336–342 (2007).
4. Emma, P. *et al. Nature Photon.* **4,** 641–647 (2010).
5. Ishikawa, T. *et al. Nature Photon.* **6,** 540–544 (2012).
6. Rohringer, N. *et al. Nature* **481,** 488–491 (2012).
7. Duguay, M. A. & Rentzepis, P. M. *Appl. Phys. Lett.* **10,** 350–352 (1967).
8. Ritson, D. M. *Nature* **328,** 487–490 (1987).
9. Matthews, D. L. *et al. Phys. Rev. Lett.* **54,** 110–113 (1985).
10. Suckewer, S., Skinner, C. H., Milchberg, H., Keane, C. & Voorhees, D. *Phys. Rev. Lett.* **55,** 1753–1756 (1985).
11. Weninger, C. *et al. Phys. Rev. Lett.* **111,** 233902 (2013).
12. Alessi, D. *et al. Phys. Rev. X* **1,** 021023 (2011).
13. Reagan, B. A. *et al. Phys. Rev. A* **89,** 053820 (2014).
14. Popmintchev, T. *et al. Science* **336,** 1287–1291 (2012).
15. Amann, J. *et al. Nature Photon.* **6,** 693–698 (2012).
16. Allaria, E. *et al. Nature Photon.* **7,** 913–918 (2013).

ARTICLE

Endosymbiotic origin and differential loss of eukaryotic genes

Chuan Ku[1], Shijulal Nelson-Sathi[1], Mayo Roettger[1], Filipa L. Sousa[1], Peter J. Lockhart[2], David Bryant[3], Einat Hazkani-Covo[4], James O. McInerney[5,6], Giddy Landan[7] & William F. Martin[1,8]

Chloroplasts arose from cyanobacteria, mitochondria arose from proteobacteria. Both organelles have conserved their prokaryotic biochemistry, but their genomes are reduced, and most organelle proteins are encoded in the nucleus. Endosymbiotic theory posits that bacterial genes in eukaryotic genomes entered the eukaryotic lineage via organelle ancestors. It predicts episodic influx of prokaryotic genes into the eukaryotic lineage, with acquisition corresponding to endosymbiotic events. Eukaryotic genome sequences, however, increasingly implicate lateral gene transfer, both from prokaryotes to eukaryotes and among eukaryotes, as a source of gene content variation in eukaryotic genomes, which predicts continuous, lineage-specific acquisition of prokaryotic genes in divergent eukaryotic groups. Here we discriminate between these two alternatives by clustering and phylogenetic analysis of eukaryotic gene families having prokaryotic homologues. Our results indicate (1) that gene transfer from bacteria to eukaryotes is episodic, as revealed by gene distributions, and coincides with major evolutionary transitions at the origin of chloroplasts and mitochondria; (2) that gene inheritance in eukaryotes is vertical, as revealed by extensive topological comparison, sparse gene distributions stemming from differential loss; and (3) that continuous, lineage-specific lateral gene transfer, although it sometimes occurs, does not contribute to long-term gene content evolution in eukaryotic genomes.

In prokaryotes, inheritance involves recombination superimposed upon clonal growth[1] and the mechanisms of recombination are the mechanisms of lateral gene transfer (LGT): transformation, conjugation, transduction, and gene transfer agents[2–4]. These mechanisms operate unidirectionally from donor to recipient and generate pangenomes[5,6]. In eukaryotes, sexual recombination is reciprocal, prokaryotic LGT machineries are lacking, and genetics indicate inheritance to be vertical[7,8]. Well-known exceptions to the vertical pattern of eukaryote evolution occurred at the origin of chloroplasts and mitochondria, where many genes entered the eukaryotic lineage via gene transfer from endosymbionts[9–11]. More controversial, however, are mounting claims for abundant and continuous LGT from prokaryotes to eukaryotes[12–17]. Such claims, if true, predict that cumulative effects of LGT in eukaryote genome evolution should be detectable in genome-wide surveys spanning many lineages. By contrast, endosymbiotic theory predicts that gene acquisitions in eukaryotes should correspond to the origins of chloroplasts and mitochondria[9] and to secondary endosymbiotic events among algae[18,19].

The evidence behind claims for widespread LGT from prokaryotes to eukaryotes, or from eukaryotes to eukaryotes, comes from genome sequences and rests upon observations of unexpected branches in phylogenetic trees[13,16] and patchy gene distributions across lineages[20,21]. Yet the same observations can stem from vertical evolution, with factors that influence phylogenetic inference causing unexpected branching patterns[22–25] and differential loss causing patchy distributions[26,27]. Distinguishing between these alternatives is not simple. Some cases of lineage-specific LGTs to eukaryotes are incontrovertible, in particular bacterial endosymbiont genome insertions into insect chromosomes[28,29] or viral acquisitions in placental evolution[30]. Yet if LGT to eukaryotes is continuously ongoing in evolution, it has to generate cumulative effects. Even if the average LGT frequency per genome is low, perhaps ~0.5% of all genes per genome[20], LGTs will still accumulate over time, like interest on a bank account: acquired genes will be inherited to descendant lineages, which themselves will continue to acquire new genes. The cumulative effect of LGT generates lineages that have increasingly different and continuously diverging collections of genes. This is exactly what is observed in prokaryotes, where known LGT mechanisms operate and pangenomes accrue[5,6]. Here we test the predictions of the competing alternatives to account for prokaryotic genes in eukaryotes—gradual LGT accrual versus episodic gene transfer from organelles—using gene distributions and maximum likelihood trees to uncover cumulative LGT effects.

Gene distributions bear out endosymbiotic theory

We clustered 956,053 protein sequences from 55 eukaryotes from six supergroups[31] and 6,103,025 sequences from prokaryotes (5,793,897 from 1,847 bacteria and 309,128 from 134 archaea) in a two-stage procedure. We first clustered all sequences within each domain (Supplementary Tables 1–5), then merged domain-specific clusters by a reciprocal best-cluster approach, resulting in 2,585 disjunct clusters containing sequences from at least two eukaryotes and at least five prokaryotes. For multidomain proteins, the cluster was assigned according to the most similar domain in the prokaryote–eukaryote comparison, favouring the detection of recent LGTs from prokaryotes, if they are present. The distributions of taxa for the 2,585 eukaryote–prokaryote clusters (EPCs) and for the 26,117 eukaryotic-specific clusters (ESCs) are shown in Fig. 1 and Extended Data Fig. 1a, respectively. The functional categories distributed across EPCs and ESCs are significantly different (Table 1 and Supplementary Table 6), reflecting the prokaryotic origin of core eukaryotic informational and operational genes[32], and the origin of eukaryotic-specific traits that followed the origin of mitochondria[33].

[1]Institute of Molecular Evolution, Heinrich-Heine University, 40225 Düsseldorf, Germany. [2]Institute of Fundamental Sciences, Massey University, Palmerston North 4474, New Zealand. [3]Department of Mathematics and Statistics, University of Otago, Dunedin 9054, New Zealand. [4]Department of Natural and Life Sciences, The Open University of Israel, Ra'anana 43107, Israel. [5]Department of Biology, National University of Ireland, Maynooth, County Kildare, Ireland. [6]Michael Smith Building, The University of Manchester, Oxford Rd, Manchester M13 9PL, UK. [7]Genomic Microbiology Group, Institute of Microbiology, Christian-Albrechts-University of Kiel, 24118 Kiel, Germany. [8]Instituto de Tecnología Química e Biológica, Universidade Nova de Lisboa, 2780-157 Oeiras, Portugal.
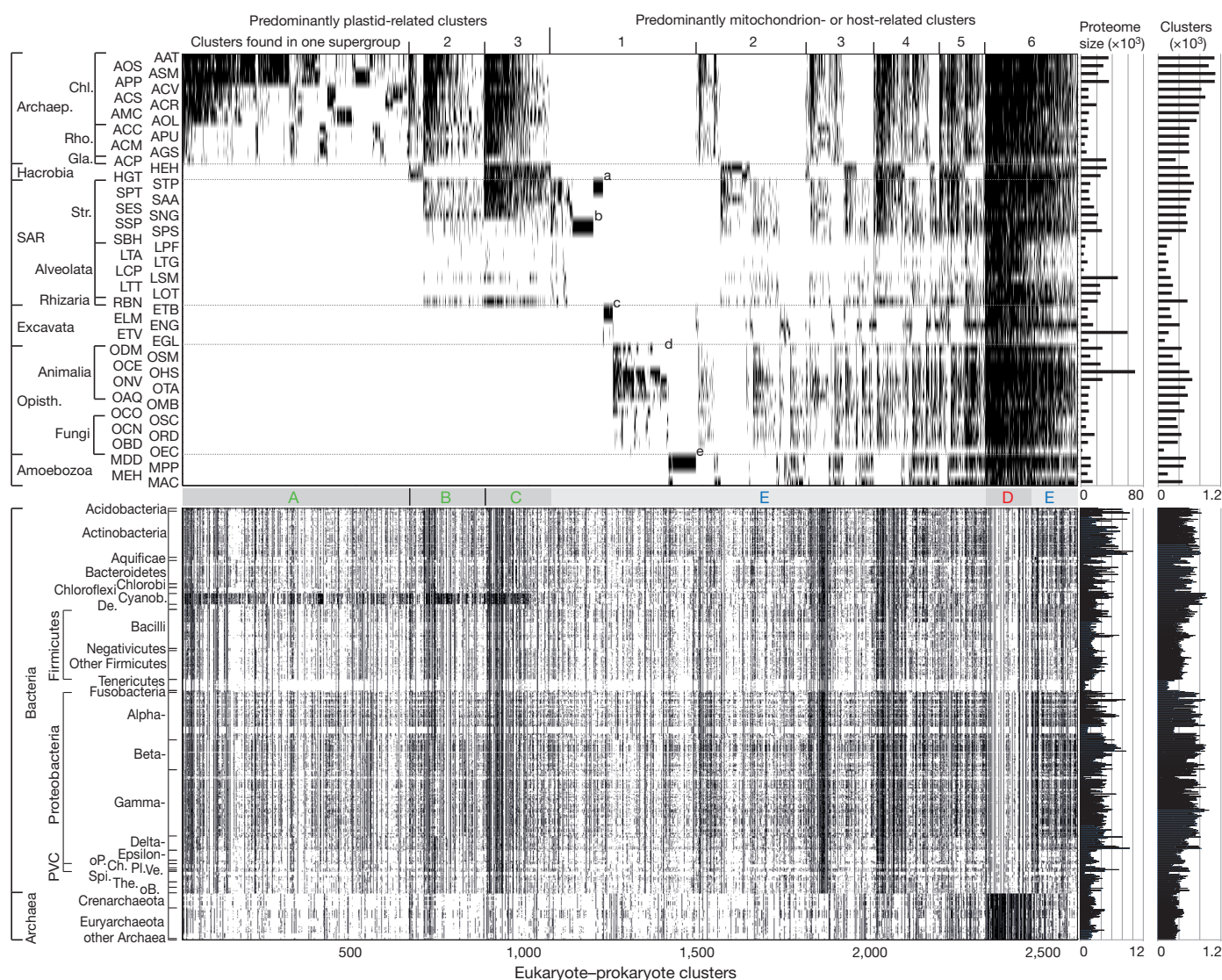
Figure 1 | Distribution of taxa in EPCs. Each black tick indicates gene presence in a taxon. The 2,585 EPCs (*x* axis) are ordered first according to their distribution across six eukaryotic supergroups with clusters specific to lineages with photosynthetic eukaryotes (blocks A–C) on the left, then according to the number of supergroups within which the clusters occur. Clusters most densely distributed in archaea among prokaryotes (block D) and others (block E) are indicated. Lower-case letters label clusters whose distribution is suggestive of recent lineage-specific acquisitions. The numbers of protein sequences and EPCs per genome are shown on the right. Taxon abbreviations are given in Supplementary Tables 1 and 3.

The phyletic distributions of the EPCs reveal blocks of genes with distinctly shared patterns that carry the unmistakable imprint of endosymbiosis in eukaryote evolution. The eukaryotic genes in blocks A–C are present in photosynthetic eukaryotes and related lineages only (Fig. 1), and are densely distributed among one particular group of prokaryotes—the cyanobacteria—as endosymbiotic theory[11] would predict. Block D encompasses genes that were present in the eukaryotic ancestor, that are very densely distributed in archaea, and that

are also more refractory to loss than any other group of eukaryotic genes. These correspond to the informational genes[32] representing the archaeal host lineage that acquired the mitochondrion in endosymbiotic theory[34–36]. The archaeal genes in eukaryotes are rarely lost (Fig. 1), being more essential than operational genes[37] and involved in information processing: unlike genes in metabolic pathways, their function cannot be replaced by importing amino acids or vitamins from the environment[29,38]. Block E encompasses many genes that

Table 1 | Functional classification of eukaryotic protein clusters

Functional category	ESCs	EPCs	EPC blocks					
			A	B	C	ABC	D	E
Cellular processes and signalling*	6,685	191	42	14	21	77	14	100
Information storage and processing*	3,940	351	67	28	27	122	75	154
Metabolism*	4,882	1,130	217	95	79	391	35	704
Poorly characterized	10,610	913	328	81	61	470	4	439
Total	26,117	2,585	654	218	188	1,060	128	1,397

The full list of clusters and functional categories is given in Supplementary Table 6. See Extended Data Fig. 10 and Methods for distribution of ESCs and EPCs under different clustering criteria and the tests comparing them.
*χ^2 test of the distribution of clusters across the three general functional categories (null hypothesis was that the distribution is independent of the sets of clusters). The sets of clusters compared (*P* value) were as follows: ESCs/EPCs (0.00), ABC/D (0.00), ABC/E (0.01), D/E (0.00), A/B (0.71), A/C (0.56), B/C (0.29).

were present in the eukaryotic common ancestor, as well as many that are shared across supergroups but are more sparsely distributed than the host-derived genes in block D. These could correspond to the mitochondrion alone[39] or to the mitochondrion plus additional donors that exist in various formulations of endosymbiotic theory[11].

Eukaryote gene distributions and origins

Among the 2,585 trees (Supplementary Table 7) plotted in Fig. 1, 1,933 (74.8%) recovered the eukaryotes as monophyletic and another 329 trees (12%) did not reject eukaryote monophyly in the Kishino–Hasegawa approximately unbiased test (AUT) (Extended Data Fig. 1b). The remaining 323 trees (12%) reject eukaryote monophyly at $P = 0.05$ in the AUT. But these 323 cases are not all necessarily bona fide cases of LGT, because endosymbiosis introduces gene redundancy (for example organelle and cytosolic ribosomes) into the eukaryotic lineage, because many sequencing contaminations are evident in these 323 trees, and because molecular phylogenetics sometimes simply fails[22–25] (Extended Data Figs 2 and 3, Supplementary Table 6 and Methods). Yet even if we assume that these 323 trees represent outright LGTs, the eukaryotes harbouring these genes are not expanding their gene content repertoire via LGT, they are merely re-acquiring members of EPC families already present in the eukaryotic lineage. Rather than dwelling on non-monophyletic exceptions, we investigated the monophyletic majority.

For the 1,933 trees that recovered eukaryote monophyly, we asked which prokaryotic groups were present in the sister group to the eukaryotic clade. Blocks A–C (Fig. 1) encompass 1,060 clusters that clearly correspond to the introduction of photosynthesis into the eukaryotic lineage[18] and its spread via secondary symbiosis[19]. The 188 genes in block C include those acquired during the cyanobacterial origin of plastids and transferred to the nucleus, and then transferred again in at least two independent secondary symbiotic events[18,19] involving the origin of (1) red secondary plastids (*Guillardia*, *Emiliania*, stramenopiles, and alveolates) and (2) green secondary plastids in the *Bigelowiella* lineage. The 218 genes in block B encompass plastid-related functions shared by Archaeplastida and one of the supergroups with secondary plastids.

The distributions of genes depicted in Fig. 1 reflect the endosymbiotic heritage of plastids far more clearly than do the underlying phylogenetic trees (Extended Data Fig. 4). Among the 889 eukaryote monophyly trees in blocks A–C (1,060 clusters), only 283 (31.8%) identified a sister group that contained cyanobacterial sequences only, while 5.9% identified a mixed sister group containing sequences from cyanobacteria and other prokaryotic groups. For the 1,397 genes in block E, 940 trees recovered eukaryote monophyly but only 5.6% identified an alphaproteobacterial sister group to eukaryotes, while 17.2% identified a mixed sister group containing sequences from alphaproteobacteria and other prokaryotic lineages. Did Archaeplastida acquire ~68% of their lineage-specific EPCs from hundreds of independent non-cyanobacterial donors, with similar, more radical implications (~94%) for the more ancient origin of the mitochondrion? That is what the trees imply, while the gene distributions suggest two episodic acquisitions, one endosymbiont donation each at the origin of plastids and mitochondria, respectively. Are the trees to be believed, or are they positively misleading? Within the EPC trees, both the prokaryote subtrees and the eukaryote subtrees address that question.

Organelle ancestors, LGT, and pangenomes

Within the prokaryotic subtrees among 2,585 EPC trees, only five prokaryotic groups were monophyletic in at least 50% of their clusters; they had no more than 15 members each. Eight prokaryotic groups were monophyletic in no more than 20% of their clusters, including alphaproteobacteria (Extended Data Fig. 2c). The extent of prokaryote non-monophyly probably reflects prokaryotic pangenomes and LGT[1–6,40]. Were eukaryotes engaging in pangenomic

LGT with prokaryotes, they would have a prokaryote-like pangenome. The 55 eukaryotic genomes sampled identify homologues in only 2,585 prokaryotic clusters. But using the same clustering criteria, 54 strains of *Escherichia coli* identify 5,074 homologous prokaryotic clusters, while samples of 55 genomes from Rhizobiales (alphaproteobacteria) recover on average 8,154 homologous prokaryotic clusters (Extended Data Fig. 2d). That is, a single bacterial species pangenome (conspecific strains of *E. coli*) has sampled prokaryote gene diversity twofold more broadly than 55 eukaryotes have in >1.5 billion years of evolution[41]. Except at organelle origins, eukaryotes are clearly isolated from the pangenome-generating LGT that prokaryotes undertake with each other, an insight that requires simultaneously investigating both phylogenies (Extended data Fig. 2c) and gene distributions (Extended data Fig. 2d).

Prokaryote pangenomes and LGT also affect the inference of gene donors to eukaryotes, because prokaryotic membership in the sister groups to eukaryotes is heterogeneous, often containing representatives from various prokaryotic phyla (Extended Data Fig. 5). Moreover, even in trees where eukaryotes branch with a sister group consisting purely of cyanobacterial, alphaproteobacterial or archaeal sequences, the eukaryotes do not branch with the same cyanobacterial, alphaproteobacterial, or archaeal sister genomes; rather, they branch with homologues from diverse members of these three prokaryotic groups (Extended Data Fig. 6). The prokaryotic homologues of genes that eukaryotes sequestered at organelle origins have been affected by pangenomes and LGT during prokaryotic genome evolution.

This effect is particularly evident in Fig. 2, where for each prokaryotic taxon the frequency of occurrence in the eukaryotic sister group is plotted against the proteome size. Only cyanobacteria, alphaproteobacteria, and, at lower significance levels, two groups of the archaea are implicated as gene donors more often than expected from random distributions of leaves in the individual trees (Supplementary Table 8). The cyanobacterial signal for plastids[11], the alphaproteobacterial
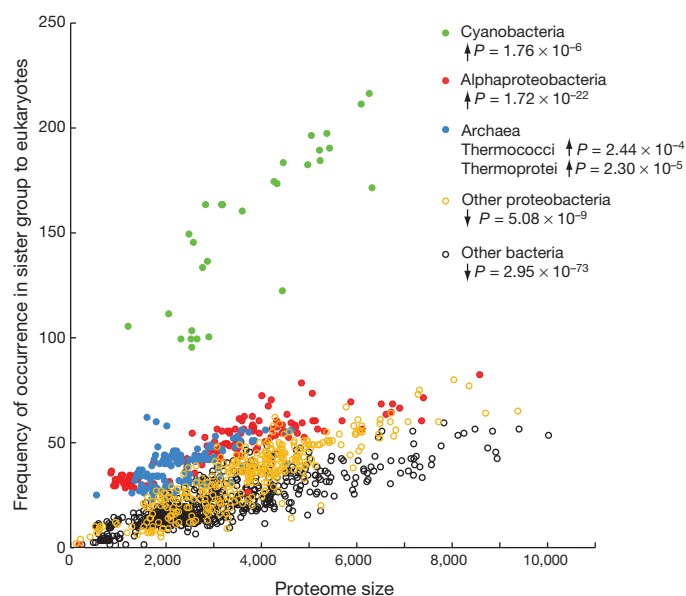

Figure 2 | Occurrence in the sister group versus proteome size. Prokaryotic taxa are plotted according to how frequently they are found in the sister group (defined as the nearest neighbour group) to a monophyletic group of eukaryotes in 1,933 trees against their proteome size. A two-sided Wilcoxon signed-rank test compares these frequencies with those generated by randomly selecting prokaryotic operational taxonomic units (OTUs) into the sister group (100 replicates). Upward and downward arrows indicate higher and lower frequencies in the real data set than in the randomized version, respectively. The test was adjusted for multiple comparisons. For complete statistics, see Supplementary Table 8.

signal for mitochondria[39], and the archaeal signal for the host[34–36] bear out the predictions of endosymbiotic theory. But beyond those three signals, no significant contributions are detected from other prokaryotes that are discussed in various formulations of endosymbiotic theory[14,42,43]. Moreover, individual trees contain information about the provenance of eukaryotic genes that is not better than random: if individual trees linking eukaryotes to prokaryotes are considered outside the context of the full set of trees to which they belong, they can—and do—deliver positively misleading results[44] about the prokaryotic subtree within which eukaryotes branch.

Eukaryote gene evolution is vertical

The eukaryote subtrees address the LGT versus endosymbiosis question even more decisively. There are only two biological mechanisms that could generate the 1,933 cases of eukaryote monophyly for the EPCs. Either the gene was present in the common ancestor of the eukaryotes possessing it and vertically inherited to descendant members[27], or it was acquired by one member of the group and then subsequently distributed via eukaryote-to-eukaryote LGT[21,45]. In the former case, the gene tree of the EPC will tend to be compatible with that observed for ESCs spanning the same taxa, whereas in the latter case the phylogenies will be very different and will differ again for each newly acquired EPC. We tested whether the ESC and EPC trees are drawn from the same distribution by comparing the similarity of trees spanning non-identical leaf sets.

Eukaryote gene evolution is resoundingly vertical (Fig. 3 and Extended Data Fig. 7), with all supergroups, and eukaryotes as a group, passing the test as not significantly different from vertical, while the eukaryote-to-eukaryote LGT alternative—a minimum topology perturbation of one random prune-and-graft LGT per tree—is strongly rejected in all cases. The crucial test case is Archaeplastida, which harbour the most supergroup-specific EPCs (Fig. 1). Although only a minority of Archaeplastida-specific EPCs phylogenetically trace directly to cyanobacteria sampled, they all trace to the Archaeplastida common ancestor (Fig. 3). The data thus indicate that (1) the Archaeplastida-specific EPCs were present in the Archaeplastida common ancestor, (2) their origin thus coincides with the origin of plastids, (3) many are directly involved in photosynthetic functions (Supplementary Table 6), but (4) the sister groups have heterogeneous membership (Extended Data Fig. 6).

This presents two alternatives. If we equate sister-group taxon labels in trees with biological donors, then plastid origin involved hundreds of independent gene donations by hundreds of different donors—the minority of them cyanobacteria—to construct, gene-by-gene, a photosynthetic eukaryote, without any of the individual donations being inactivated through mutation before the plastid was assembled to a functional unit. Alternatively, the gene trees are positively misleading, and these Archaeplastida-specific EPCs were acquired from the ancestor of plastids, which had a fully functional photosynthetic apparatus that merely needed to be integrated into the eukaryotic lineage via recurrent transfer of the necessary genes from the resident organelle to the nucleus[9], clearly the preferable alternative. The untenable proposition of gene-by-gene plastid assembly via hundreds of targeted LGTs arises from interpreting the trees, which can be positively misleading, at face value.

Episodic influx and differential loss

The Archaeplastida case is so important because exactly the same set of observations and the same reasoning applies to the mitochondrion. The host for the origin of plastids was a heterotroph; the transition to autotrophy was driven by endosymbiosis and gene transfer[9,11]. The gene distributions (Fig. 1) reflect that. Similarly, the host for the origin of mitochondria was an archaeon[34–36], the transition to chemiosmotic ATP synthesis in the mitochondrion also resulting from endosymbiosis and gene transfer from the organelle to the host[33]. As with plastids, mitochondria cannot have been constructed via one-by-
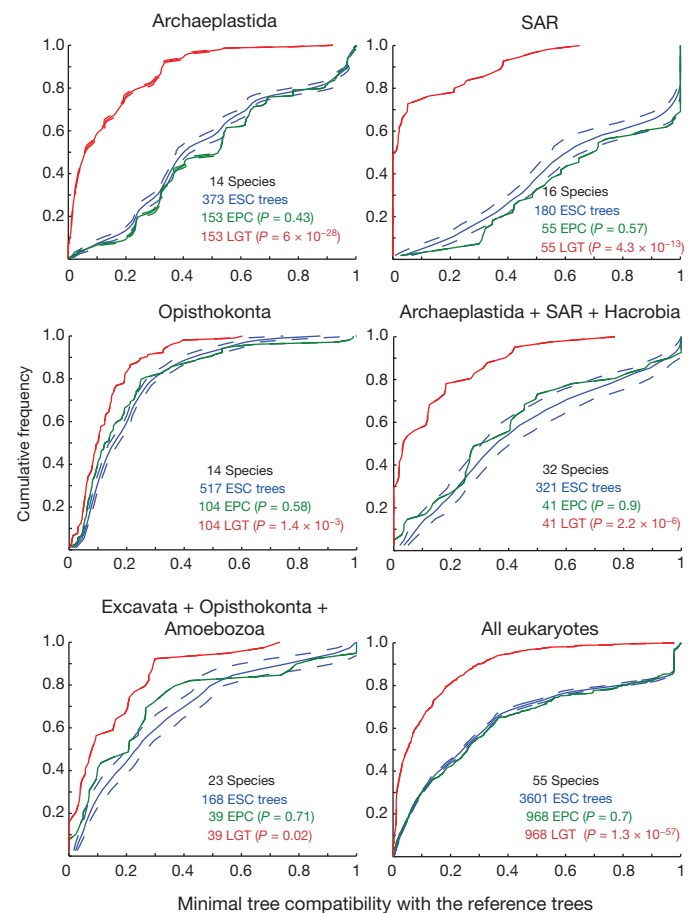

Figure 3 | Comparison of sets of trees for single-copy genes in eukaryotic groups. Cumulative distribution functions (y axis) for scores of minimal tree compatibility with the vertical reference data set (x axis). Values are number of species, sample sizes, and P values of the two-tailed Kolmogorov–Smirnov two-sample goodness-of-fit test in the comparison of the ESC (blue) data sets against the EPC (green) data set and a synthetic data set simulating one LGT (red). Dashed lines delineate the range of distributions in 100 replicates of random down-sampling. See also Extended Data Fig. 7.

one LGT, because hundreds of randomly acquired genes to assemble a respiratory organelle cannot be maintained by purifying selection until the mitochondrion is fully functional. Gene transfer from a respiring endosymbiont[9,46] is, by comparison, facile.

Vertical gene inheritance in eukaryotes (Fig. 3 and Extended Data Fig. 7) has a further consequence: the patchy distributions of genes across eukaryotic lineages sampled here are not the result of eukaryote-to-eukaryote LGT, they are the result of differential loss. This is true not only for the EPCs shown in Fig. 1 but also for the ESCs (Extended Data Fig. 1a). Patchy gene distributions in prokaryotes generally indicate LGT, except in isolated species undergoing reductive evolution[38]. In eukaryotes, patchy distributions are often interpreted as evidence for LGT[13], yet the present findings show that patchy distributions in eukaryotes are better explained by differential loss. This leads to steadily declining genome size in terms of numbers of EPCs across eukaryote phylogeny (Extended Data Fig. 8a), with the notable exception of the origin of Archaeplastida, where EPCs double by the influx of ~1,000 clusters. Gene acquisitions in eukaryotes are episodic and correspond to symbioses (Extended Data Fig. 8b).

Finally, some gene distributions among EPCs are highly suggestive of lineage-specific acquisition, because many lineage-specific losses must be assumed. These include 67 dictyostelid-specific genes and 160 opisthokont-specific genes directly observable in Fig. 1, and 210 genes putatively acquired by the ancestor of land plants (Extended Data Fig. 9a). Were these genes recent LGTs, for example during land
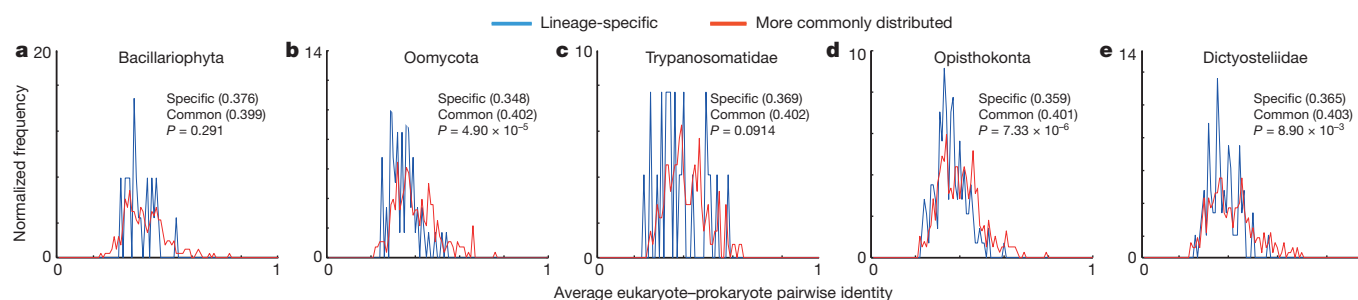
Figure 4 | Eukaryote–prokaryote sequence identities for genes with a tip distribution in eukaryotes versus those whose distributions trace their presence to a more ancient ancestor. a–e, Genes denoted by lower-case letters in Fig. 1 and those found in at least three of five major supergroups. The mean of the average pairwise identities is shown in parentheses. At $P = 0.05$, a two-sided Wilcoxon rank-sum test either did not reject the null hypotheses that the two sets of genes are not different (**a**, **c**) or suggested the tip-specific eukaryotic genes are less similar to their prokaryotic homologues (**b**, **d**, **e**). See also Extended Data Fig. 9.

plant origin ~450 million years ago[47], they should be more similar to their prokaryotic sisters than genes acquired at plastid and mitochondrial origin. The converse is observed (Fig. 4 and Extended Data Fig. 9). While we do detect genome-specific candidate LGTs (cLGTs), namely eukaryotic singletons that show high similarity to prokaryotic genes, their frequency is approximately four to ten times lower than that of nuclear insertions of mitochondrial and chloroplast DNA[46] (Supplementary Table 9). Thus, even on short timescales, the contribution of gene transfers from organelles is greater than that of cLGTs, whose numbers tend to decrease with updated genome annotations.

Conclusion

Eukaryote gene content evolution resembles the situation in archaea, where gene transfer also has an episodic tendency[48]. Despite many reports of LGT to and among eukaryotes, the combined analyses of all trees that would address the issue reveal no evidence for a detectable cumulative impact of continuous LGT on the evolution of eukaryote gene content. This indicates either (1) that lineage-specific LGTs rapidly undergo loss, having short residence times within their corresponding lineages, (2) that LGT-prone lineages do not give rise to evolutionarily stable descendants, with LGTs being concentrated in evolutionary dead-ends in a kind of terminal differentiation[49], (3) that many suspected LGTs are not really lineage-specific after all and with further eukaryote sampling they will eventually crop up in other distantly related eukaryotes as evidence for differential loss, or (4) any combination thereof. Eukaryotes obtain novel gene families via gene and genome duplication, prokaryotes undergo LGT[50]. Two episodes of gene influx—one from mitochondria and one from chloroplasts, followed by differential loss—account for the phylogeny and distribution of bacterial genes in eukaryotes, which sampled prokaryotic pangenomes at organelle origins.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55,** 709–742 (2001).
2. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284,** 2124–2128 (1999).
3. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).
4. Lang, A. S., Zhaxybayeva, O. & Beatty, J. T. Gene transfer agents: phage-like elements of genetic exchange. *Nature Rev. Microbiol.* **10,** 472–482 (2012).
5. Rasko, D. A. *et al.* The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **190,** 6881–6893 (2008).
6. Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Gene frequency distributions reject a neutral model of genome evolution. *Genome Biol. Evol.* **5,** 233–242 (2013).
7. Szathmáry, E. & Maynard Smith, J. The major evolutionary transitions. *Nature* **374,** 227–232 (1995).
8. Nei, M. *Mutation-Driven Evolution* (Oxford Univ. Press, 2013).
9. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5,** 123–135 (2004).
10. Lane, C. E. & Archibald, J. M. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol. Evol.* **23,** 268–275 (2008).
11. Archibald, J. M. *One plus One Equals One: Symbiosis and the Evolution of Complex Life* (Oxford Univ. Press, 2014).
12. Andersson, J. O. Lateral gene transfer in eukaryotes. *Cell. Mol. Life Sci.* **62,** 1182–1197 (2005).
13. Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nature Rev. Genet.* **9,** 605–618 (2008).
14. Price, D. C. *et al. Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335,** 843–847 (2012).
15. Boto, L. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc. R. Soc. B* **281,** 20132450 (2014).
16. Huang, J. L. Horizontal gene transfer in eukaryotes: the weak-link model. *Bioessays* **35,** 868–875 (2013).
17. Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16,** 50 (2015).
18. Gould, S. B., Waller, R. R. & McFadden, G. I. Plastid evolution. *Annu. Rev. Plant Biol.* **59,** 491–517 (2008).
19. Curtis, B. A. *et al.* Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492,** 59–65 (2012).
20. Alsmark, C. *et al.* Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biol.* **14,** R19 (2013).
21. Keeling, P. J. & Inagaki, Y. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. *Proc. Natl Acad. Sci. USA* **101,** 15380–15385 (2004).
22. Steel, M., Penny, D. & Lockhart, P. J. Confidence in evolutionary trees from biological sequence data. *Nature* **364,** 440–442 (1993).
23. Lockhart, P. J. *et al.* A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* **15,** 1183–1188 (1998).
24. Guo, Z. H. & Stiller, J. W. Comparative genomics and evolution of proteins associated with RNA polymerase II C-terminal domain. *Mol. Biol. Evol.* **22,** 2166–2178 (2005).
25. Semple, C. & Steel, M. *Phylogenetics* (Oxford Univ. Press, 2003).
26. Hughes, A. L. & Friedman, R. Loss of ancestral genes in the genomic evolution of *Ciona intestinalis. Evol. Dev.* **7,** 196–200 (2005).
27. Müller, M. *et al.* Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76,** 444–495 (2012).
28. Kondo, N., Nikoh, N., Ijichi, N., Shimada, M. & Fukatsu, T. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc. Natl Acad. Sci. USA* **99,** 14280–14285 (2002).
29. Husnik, F. *et al.* Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153,** 1567–1578 (2013).
30. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403,** 785–789 (2000).
31. Derelle, R. *et al.* Bacterial proteins pinpoint a single eukaryotic root. *Proc. Natl Acad. Sci. USA* **112,** E693–E699 (2015).
32. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95,** 6239–6244 (1998).
33. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467,** 929–934 (2010).
34. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504,** 231–236 (2013).
35. Guy, L., Saw, J. H. & Ettema, T. J. G. The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6,** a016022 (2014).
36. Koonin, E. V. & Yutin, N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6,** a016188 (2014).

37. Cotton, J. A. & McInerney, J. O. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc. Natl Acad. Sci. USA* **107,** 17252–17255 (2010).
38. Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42,** 165–190 (2008).
39. John, P. & Whatley, F. R. *Paracoccus denitrificans* and the evolutionary origin of the mitochondrion. *Nature* **254,** 495–498 (1975).
40. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36,** 6688–6719 (2008).
41. Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108,** 13624–13629 (2011).
42. Margulis, L., Dolan, M. F. & Guerrero, R. The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc. Natl Acad. Sci. USA* **97,** 6954–6959 (2000).
43. Fuerst, J. A. & Sagulenko, E. Keys to eukaryality: Planctomycetes and ancestral evolution of cellular complexity. *Front. Microbiol.* **3,** 167 (2012).
44. Domman, D., Horn, M., Embley, T. M. & Williams, T. A. Plastid establishment did not require a chlamydial partner. *Nature Commun.* **6,** 6421 (2015).
45. Hug, L. A., Stechmann, A. & Roger, A. J. Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. *Mol. Biol. Evol.* **27,** 311–324 (2010).
46. Kleine, T., Maier, U. G. & Leister, D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu. Rev. Plant Biol.* **60,** 115–138 (2009).
47. Yue, J. P., Hu, X. Y., Sun, H., Yang, Y. P. & Huang, J. L. Widespread impact of horizontal gene transfer on plant colonization of land. *Nature Commun.* **3,** 1152 (2012).
48. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *Bioessays* **35,** 829–837 (2013).
49. Hao, W. L. & Golding, G. B. The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res.* **16,** 636–643 (2006).
50. Treangen, T. J. & Rocha, E. P. C. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* **7,** e1001284 (2011).

Supplementary Information is available in the online version of the paper.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Sequence clustering. Protein sequences were downloaded from the NCBI database (version June 2012) for complete prokaryotic genomes and from respective genome sequencing websites for a phylogenetically diverse range of eukaryotes (Supplementary Table 1). Eukaryotic, bacterial, and archaeal protein sequences were clustered separately before homologous clusters from eukaryotes and prokaryotes were combined. The bacterial sequences (Supplementary Table 3) and the archaeal sequences (Supplementary Table 4) were clustered using the methods described[51] ('needle' global identity ≥25%). Eukaryotic sequences were clustered with the reciprocal best BLAST[52] (version 2.2.28; cut-off: expect (E) value ≤1×10^{-10}) hit (rBBH) procedure[53] followed by calculation of pairwise global identity (cut-off: global identity ≥40%) of each rBBH pair using the program 'needle' in the EMBOSS package[54] and MCL clustering[55] on the basis of the global identities. Because the prokaryotic genome sample is biased towards bacteria and because many bacterial species are represented by multiple strains (up to 54 for *E. coli*), before clustering, genome sequences from bacterial strains were combined into species pangenomes (Supplementary Table 3) and the rBBH procedure for bacteria (cut-off: E value ≤1×10^{-10} and local identity ≥30%) was performed at the species level to take overrepresentation of bacteria and heavily sequenced bacterial species into account. To avoid combining clusters with different homologous protein domains due to gene fusion or recombination[56], a reciprocal best cluster procedure was used to compare and combine eukaryotic with prokaryotic clusters. Reciprocal all-against-all BLAST searches (cut-off: E value ≤1×10^{-10} and local identity ≥30%) were conducted between 136,661 sequences in all 28,702 eukaryotic clusters containing sequences from at least two eukaryote genomes each, and 4,154,013 sequences in 102,089 bacterial clusters as well as 232,046 sequences in 11,992 archaeal clusters. Prokaryotic clusters containing sequences from not more than four taxa (Supplementary Table 1) were excluded. If ≥50% of the sequences of a eukaryotic cluster had their best hit in a bacterial or archaeal cluster, they were designated the best bacterial or archaeal cluster of the eukaryotic cluster, and vice versa. When a eukaryote cluster and a prokaryote cluster were reciprocally the best clusters for each other, the prokaryotic cluster was combined with the eukaryotic cluster, resulting in an EPC. In total, 2,585 EPCs containing one eukaryote cluster and one bacterial, one archaeal, or two prokaryotic clusters were obtained; the 26,117 remaining eukaryotic clusters were designated ESCs.

Different sets of EPCs and ESCs were generated with lowered thresholds for identifying the best cluster, including changing the BLAST local identity cut-off from 30% to 20% and the minimum proportion of sequences having the best hit in a cluster (best-hit correspondence) from 50% to 40%, 30%, 20% and 10%. Lowering the best-hit correspondence threshold to ≤50% can generate more than one 'best' cluster. To avoid combining two 'best' clusters corresponding to different domains of the sequences in the query cluster into one EPC, we adhered to the >50% threshold. Lowering the local identity or best-hit correspondence thresholds converts some ESCs to EPCs, but the distribution of clusters across eukaryotic taxa is not changed (Extended Data Fig. 10) and the distribution of the functional categories of the clusters remains significantly different between ESCs and EPCs (Table 1; $P = 0.00$ for all thresholds in a χ^2 test). Different EPC sets generated with different thresholds are samples from the same pool of eukaryotic genes derived from prokaryotes; sampling lower thresholds for sequence conservation increases the proportion of poorly conserved genes in the alignment and phylogeny steps.

Functional annotation and test of independence. All eukaryotic protein sequences from the 28,702 clusters were BLASTed (cut-off: E value ≤1×10^{-10} and local sequence identity ≥50%) against the eggNOG version 4.0 (ref. 57) database, and the eggNOG/cluster of orthologous groups (COG) identifier of the best hit was assigned to each sequence. A particular eggNOG/COG identifier was assigned to a cluster if it was assigned to more sequences in that cluster than any other identifier. Ties were broken by taking the first listed identifier. Each identifier was then mapped to the COG functional categories[58]. If an identifier was mapped to two or more categories, the category R (general function prediction only) was assigned. Functional annotations are in Supplementary Table 6.

If two sets of eukaryotic genes originated from different prokaryotic sources, the distribution of the functional categories should reflect that of the sources and could be significantly different. To test this, the COG functional categories were divided into four major categories: cellular processes and signalling, information storage and processing, metabolism, and poorly categorized proteins (including those clusters not assigned any eggNOG/COG identifier). A χ^2 test of independence (Table 1) was then used to compare the distribution of genes in the three

former categories between ESCs and EPCs (on the basis of different thresholds for combining eukaryote and prokaryote clusters) and between the different blocks of EPCs (Fig. 1) that mainly corresponded to different sources (ABC, D, E) or the same one (A, B, and C).

Relationships between subgroupings within eukaryotes, archaea, and bacteria. A backbone tree of eukaryotes was constructed on the basis of recently published phylogenies[31,59–68]. The archaeal tree was based on the 70 single-copy genes present in the archaeal clusters and was generated in a previous study[51]. Since there was no single-copy orthologue present in every bacterial taxon, 32 nearly universal (present in at least 1,780 out of the 1,847 genomes) single-copy genes were used for inference of a bacterial reference tree (Supplementary Table 3). The OTU for the tree was species (see above). When a species pangenome had multiple sequences (in most cases, each from a different strain of the species) in a cluster, the first in the sorted list of the NCBI GI numbers was used as the representative sequence for this species. The sequences from each gene were aligned separately using MAFFT version 7.130 (ref. 69) with the option 'linsi' and concatenated into a single alignment. A maximum likelihood tree was reconstructed using RAxML version 7.8.6 (ref. 70) under the PROTCATWAG model. An initial tree revealed that some species had much longer branches. A second RAxML run was conducted without four long-branch taxa ('*Candidatus* Tremblaya princeps', '*Candidatus* Hodgkinia cicadicola', '*Candidatus* Zinderia insecticola', and '*Candidatus* Carsonella ruddii'). The reference tree generated was used to modify the taxonomic assignment of some taxa. For example, according to NCBI Taxonomy, *Erysipelothrix rhusiopathiae* strain Fujisawa is placed under Firmicutes in its own class, but the reference tree shows that it is nested within the clade formed by Tenericutes, so it should be placed under this phylum (as is also suggested by a recent study[71]). The curated taxonomic information for bacteria can be found in Supplementary Table 3.

Alignment, phylogenetic analyses, and test for eukaryote monophyly. Sequences in each of the 2,585 EPCs were aligned using MAFFT version 7.130 (ref. 69) with the option 'linsi'. The quality of alignment was compared between different sets of clusters using the HoT method[72,73] with the programs COS_v2.05.pl (in combination with MAFFT 7.130) and msa_set_score_v2.02. Maximum likelihood trees were reconstructed using RAxML version 7.8.6 (ref. 70) under the PROTCATWAG model, with special amino-acid characters U and J converted to X (unknown). The trees (Supplementary Table 7) were analysed using custom Perl scripts to determine whether the eukaryotic sequences formed a clade (Supplementary Table 6); if they did, the prokaryotic clade with the smaller average distance to the eukaryotic clade was identified as the sister group. This criterion is favoured over the use of the number of taxa in the neighbouring groups because the different prokaryotic higher-level taxonomic groups vary greatly in the number of species and genomes sampled (Supplementary Tables 3 and 4).

In cases where the eukaryotic sequences did not form a clade, we conducted the AUT implemented in the CONSEL package[74] to determine whether the apparent non-monophyly was statistically significant. From the maximum likelihood tree of each of the 652 EPCs where eukaryotes were recovered as non-monophyletic, we extracted a eukaryotic subtree by pruning the prokaryotic sequences and a prokaryotic subtree by pruning the eukaryotic sequences. We then generated the set of all trees formed by re-grafting the subtree with eukaryotic sequences onto the subtree of prokaryotic sequences, keeping those closest to the original maximum likelihood tree in terms of Robinson and Foulds[75] distance (as computed by the program treedist of the PHYLIP package[76] version 3.695). For all these candidate trees, PhyML version 3.1 (ref. 77) was used to optimize parameters and calculate per-site likelihoods, using option –print_site_lnl, the WAG[78] evolution model, 25 evolutionary rate categories, estimation of gamma distribution shape parameter alpha, and by providing the alternative tree(s) as user tree. Note that only branch lengths and rate parameters, but not topology, were optimized using the -o lr option.

The program makermt in CONSEL version 1.16 was used with –phyml option and a file containing the site-likelihoods for the original tree together with those for the alternatives as input. The output file of makermt was provided to CONSEL version 1.20 and the program catpv used to extract P values for the tree set.

If at least one of the alternative trees lay in the confidence interval of the original tree, namely in cases where the P value of the AUT from the multiple bootstrap (au) was not significant at the 5% level, the alternative tree with monophyletic eukaryotic sequences was considered to be equally likely (that is, not significantly worse than the original tree) and monophyly was not rejected (Extended Data Fig. 1b and Supplementary Table 6).

Classification of eukaryote non-monophyly trees. The 323 EPCs that failed the AUT for eukaryote monophyly were manually inspected and classified into categories according to the distribution of eukaryotic sequences in the respective phylogenetic trees. The categories were assigned as follows. Eukaryotes appear as one clade with the exception of sequences from at most one (1) or

two (2) eukaryotes as outlier(s). (3) Eukaryotes appear in two separate clades. Archaeplastida/SAR (stramenopiles + alveolates + Rhizaria)/Hacrobia (photosynthetic eukaryotes and their relatives) and the other eukaryotes form two separate clades (4) with the exception of sequences from at most one eukaryotic outlier (5). Cyanobacterial sequences branch within a single clade of Archaeplastida/SAR/Hacrobia (6) with the exception of one (7) or two (8) eukaryotic outlier(s). (9) Trees contain sequences from only two distinct eukaryotes that do not form a clade. (10) Trees where eukaryotic monophyly could be achieved by removing one sequence or one small clade of prokaryotes. (11) Remaining trees with more complex interleaving of prokaryotic and eukaryotic sequences. The frequency of outlier organisms in the trees was recorded (Supplementary Table 6). To investigate the relationship of gene-copy numbers with eukaryotic monophyly within EPCs, the number of EPCs containing more than one sequence per eukaryote was counted. A χ^2 goodness-of-fit test was used to compare different categories of EPCs with the eukaryote monophyletic EPCs; significance values at the 5% level are reported (Supplementary Table 6).

Prokaryotic gene sharing by eukaryotes and prokaryotes. To compare the number of genes shared by eukaryotes and prokaryotes and those by prokaryotic groups and other prokaryotes, we performed the same clustering procedure as used to generate EPCs for the prokaryotic groups shown in Fig. 1. Protein sequences from 55 prokaryote genomes randomly sampled from within a given group were clustered, as were sequences from the prokaryotes excluding the group, using the same criteria as those used to generate EPCs. The clusters from the sample were combined with the other clusters using the reciprocal best cluster procedure. The number of clusters shared between the 55-prokaryote sample and the remaining prokaryotes was counted (Extended Data Fig. 2d). The procedure was repeated for 100 random samples of 55 genomes (or a single sample of 54 *E. coli* genomes in our data set). Relative to eukaryotes, the extent of prokaryote gene sharing is slightly underestimated owing to smaller prokaryote gene pools as a result of removal of the given group.

Randomization test. All prokaryotic higher-level taxa and almost all prokaryotic species sampled occur in the sister group to eukaryotes in at least one tree (Supplementary Table 8); but instead of bona fide direct gene transfers to eukaryotes, this could result from phylogenetic errors and other factors such as LGT among prokaryotes and gene loss[79]. To evaluate whether the number of times a particular group identified as a putative donor lineage was statistically significant, we compared this number with the expected number of donor inferences in randomized versions of the phylogenetic trees. The frequency of occurrence was counted as the number of trees in which any sequence from a species was found in the sister group to eukaryotes (Fig. 2). The counting was performed for the 1,933 eukaryote monophyletic trees and for 1,933 trees with the same OTUs and the sister group of the same size where OTUs were randomly chosen to be in the sister group. The randomization procedure was repeated 100 times and the counts were averaged. A two-sided Wilcoxon signed rank test was performed in MATLAB R2013a (signrank) with the null hypothesis that the frequency of occurrence normalized by the proteome size for taxa from a taxonomic group was not different between the original 1,933 trees and the randomized data set. A procedure for controlling the false discovery rate[80,81] was used to correct for multiple comparisons involving different taxonomic groups.

Comparison of tree sets. Data sets. We considered six species groupings: (1) Archaeplastida; (2) SAR; (3) Opisthokonta; (4) Archaeplastida, SAR, and Hacrobia; (5) Excavata, Opisthokonta, and Amoebozoa; and (6) any eukaryotic group(s). The data set for each grouping consisted of three tree sets: (1) the verticality reference set consisting of the ESC trees, restricted to the species under consideration; (2) the imports set consisting of the EPC trees, restricted to the species under consideration; and (3) a synthetic data set, 'LGT', derived from the EPC set (2) by the introduction of one random LGT event, simulated by a random prune-and-graft topological operation. Only trees with more than three eukaryotic taxa were considered, which were further subject to two inclusion variants: (1) trees where the gene was present as a single-copy gene in each eukaryote, and where the eukaryote taxa were monophyletic (Fig. 3); and (2) a more inclusive criterion, where intraspecific paralogues (inparalogues) in the EPC/ESC trees were reduced to one before the remaining eukaryote sequences were realigned and trees re-done, EPCs that passed the AUT for eukaryote monophyly (Supplementary Table 6) were included, and species with multiple copies of the gene were allowed (Extended Data Fig. 7). In the last case, multiple-gene-copy taxa were pruned from the tree to avoid paralogy obfuscation. ESC and EPC trees in Newick format for these two inclusion variants can be found in Supplementary Tables 1 and 7.

Congruence tests. The congruence of individual trees or sub-trees with the entire ESC tree set was measured using the minimal compatibility measure[51]. The trees in each set were layered according to the number of taxa, and pooled together using the random down-sampling procedure[51]. We performed 100 replicates of

this procedure, and for each set derived the average cumulative distribution function. The fit between the ESC reference set and the EPC imports and LGT set was tested using a two-tailed Kolmogorov–Smirnov two-sample goodness-of-fit test[82], operating on the average cumulative distributions of the minimal compatibility scores.

Code availability. The MATLAB code used to compare tree sets (Fig. 3 and Extended Data Fig. 7) is available in the source data for Fig. 3.

Identities between eukaryote sequences and prokaryote sister-group sequences. Gene families that are specific to a eukaryotic group or where it forms a distinct clade from other eukaryotes in the tree raise the possibility of a recent lineage-specific transfer. If that were the case, such genes (recent set) are expected to exhibit higher similarities to their prokaryote homologues than more ancient acquisitions (ancient set). To test this, we performed two comparisons of eukaryote–prokaryote sequence identities between the two sets of genes. In the first comparison (Fig. 4), the recent set comprised genes specific to a eukaryote lineage. These are marked with lower-case letters in Fig. 1 and include 28 genes present in bacillariophytes in Fig. 4a, 59 genes present in oomycetes in Fig. 4b, 26 genes present in trypanosomatids in Fig. 4c, 160 genes present in opisthokonts Fig. 4d, and 67 genes present in dictyostelids in Fig. 4e. The ancient set consist of genes commonly present in eukaryotes (found in at least three supergroups, excluding Hacrobia, which are too narrowly sampled). Pairwise sequence identities were calculated as the fraction of amino-acid positions identical between two sequences in the EPC alignments using the program protdist of PHYLIP[76]. For the recent set, pairwise identities were calculated for any eukaryote sequence in the respective monophyletic clade of group-specific genes (lower-case letters in Fig. 1) and all prokaryote sequence in the respective sister group. For the ancient set, pairwise identity was calculated between any sequence from the target eukaryote lineage (for example all bacillariophytes in Fig. 4a) and any prokaryote sequence in the sister group to eukaryotes, in trees where all eukaryote sequences were monophyletic.

For the second comparison (Extended Data Fig. 9), we analysed all EPC trees to test the possibility that LGT from prokaryotes occurred continuously throughout eukaryote lineages. Genes were sorted into potentially recent and potentially ancient acquisition bins. Several criteria were applied to determine whether a gene was probably acquired in a eukaryote common ancestor (for example present in Chloroplastida + Rhodophyta) on the basis of gene distribution, as follows. (1) The gene needs to have a high density distribution: present in at least 33% of the species sampled for each descendent lineage. In the example of (Chloroplastida + Rhodophyta), at least three green lineage and two red lineage members should have the gene. (2) All sequences from this lineage form a clade in the tree. (3) The sister group to this clade consists of only prokaryote sequences. The patterns suggestive of LGT within each supergroup were inferred under these criteria and mapped onto the eukaryote reference tree (Extended Data Fig. 9a). They were separated into two sets based on the age of the last common ancestor of the eukaryote lineage that apparently acquired the gene: if the last common ancestor was younger than 800 million years according to the reference time tree of eukaryotes[41], the apparent LGT belonged to the recent set; if not, it belonged to the ancient set. In total, the numbers of genes included in recent/ancient sets were 417/254 (Archaeplastida), 130/17 (SAR), 48/4 (Excavata), 41/70 (Opisthokonta), and 79/12 (Amoebozoa). If the age of a particular node (for example, the last common ancestor of *Dictyostelium* and *Polysphondylium*) could not be inferred from the reference time tree, its age was inferred on the basis of its position relative to other nodes in reference trees for the individual supergroups (for example, ref. 64). Pairwise identities were calculated between any sequence in the recipient eukaryote lineage and any prokaryote sequence in the sister group.

For both comparisons, all pairwise identities were averaged for each tree. In Fig. 4 and Extended Data Fig. 9b, the frequencies of the average pairwise identities were normalized so that the area under the curve equalled one. A two-sided Wilcoxon rank-sum test (MATLAB: ranksum) was used to compare identities between the two sets of genes.

Reductive genome evolution in eukaryotes. Our results suggest that the vast majority of EPCs originated from only three prokaryotic donors and have been vertically inherited, followed by differential loss. This is indicated by the gene distributions themselves (Fig. 1), the presence of only three significant prokaryotic donors (Fig. 2), verticality of eukaryotic genes (Fig. 3 and Extended Data Fig. 7), lack of evidence for recent acquisitions based on sequence identity (Fig. 4 and Extended Data Fig. 9), and a strong barrier against LGT between prokaryotes and eukaryotes (Extended Data Fig. 2d). Under this premise, eukaryote ancestral genome sizes were reconstructed using a loss-only model[83] by assuming that all genes in blocks D and E and in blocks A–C originated at the root of eukaryotes and the root of Archaeplastida, respectively, and that patchy distributions result from differential loss. Although it is widely accepted that secondary symbioses spread genes from green algae to two eukaryotic lineages via secondary symbiosis, the number and nature of secondary symbioses giving rise to plastids in the

Hacrobia and SAR lineages (blocks B and C in Fig. 1) is still a matter of debate[18,19,67]. Therefore, for Hacrobia and SAR, genes in blocks B and C were not counted as part of the ancestral genome size (Extended Data Fig. 8a).

Symbiosis and gene transfer in eukaryote genome evolution. Prokaryote reference trees were generated. The archaeal reference tree was condensed into a 13-OTU backbone tree, with each OTU representing a major group of archaea. RAxML trees were reconstructed using the same parameters for each individual gene of the 70 single-copy genes used for the backbone tree, with taxa from each archaeal group constrained to be monophyletic. Similarly, individual gene trees were reconstructed for the 32 bacterial genes, with taxa from each of the 23 major groups constrained to be monophyletic. The non-Bacilli and non-Negativicutes Firmicutes, which form a grade instead of a clade, were forced to be monophyletic and collectively denoted 'Clostridia'. To see how well the individual trees supported the reference tree and how their topologies conflicted with each other, each individual tree was compared with the reference tree and each branch on the latter was colour-coded by how often (white: 0%; black: 100%) the proximal node of this branch was recovered. The bacterial tree was arbitrarily rooted with Thermotogae and the archaeal root was put between Euryarchaeota and the other archaea, a position similar to a recently proposed one[84] except that Nanoarchaeota is not regarded as part of Euryarchaeota.

To indicate the distribution of the nearest prokaryotic neighbours of eukaryotic genes (Extended Data Fig. 8b), which according to the present data were mainly acquired in the eukaryote ancestor and the archaeplastidan ancestor, the prokaryote taxa in the sister group to eukaryotes were mapped with lateral edges linking prokaryotic groups to eukaryotic nodes corresponding to endosymbiotic events: the origin of mitochondria, the origin of plastids, and secondary symbioses. To avoid assigning genes to the wrong source, more conservative criteria were adopted. For the plastid origin, a gene needs to be present in at least two Archaeplastida species, the sequences from Archaeplastida need to be monophyletic or, given secondary endosymbiosis, form a clade where Hacrobia or SAR species are nested (that is, neither of the two descendent lineages of the root of this clade consists of purely Hacrobia or SAR), and the sister group to this clade needs to consist of prokaryotes instead of eukaryotes. Any prokaryotic group occurring in the sister group was counted once and a total frequency was calculated for each group across all trees. The lateral edges linking prokaryotic and eukaryotic trees were colour-coded according to the total frequencies. The reference trees used were the eukaryote reference tree and the prokaryotic backbone trees with shadings showing signal incongruence between individual genes used to construct each tree. For red secondary symbiosis, only one event is indicated for simplicity, but the single lateral red edge makes no statement about the number or timing of events that might have occurred in evolution. Similarly, two secondary symbioses involving green plastids have occurred, but plastid-bearing euglenids are not present among the current genome sample.

Recent organelle insertions in eukaryote genomes. Mitochondrial, plastid, and nuclear genomes were downloaded (Supplementary Table 1). Out of 55 genomes, given the available organelle data, we were able to analyse 39 nuclear genomes for the existence of nuclear mitochondrial DNA copies (numts) and 24 nuclear genomes for the existence of nuclear plastid DNA copies (nupts). Each organelle genome was BLASTed against the corresponding nuclear genome using Blast+[85] with the blastn task, E value $\leq 1 \times 10^{-4}$, and with dust flag on for masking low-complexity regions. With a combination of in-house Perl scripts and MySQL queries, the BLAST hits were further filtered and counted as described below. To avoid including contaminating organelle DNA sequences in the count, only BLAST hits with a subject (contig) coverage of <70% were retained. Two different sets of criteria were then applied to produce two sets of BLAST hits: hit identity $\geq 80\%$ and length ≥ 100 base pairs, or hit identity $\geq 95\%$ and length ≥ 50 base pairs. Hits by identical sequences in different positions of the organelle were counted only once. To estimate the minimal number of independent insertion events in each nuclear genome, the following approach was applied. First, when several organelle fragments had hits to the exact same nuclear fragment, one was randomly chosen. Next, if several organelle fragments had hits to overlapping nuclear fragments, the longer one was chosen for further analysis. Finally, closely spaced organelle hits were concatenated if the nuclear distance between them was smaller than 2 kilobases. This is a permissive version of the method described in ref. 86. To get a minimum estimate, we chose here to concatenate any tandem organelle hits and hits on both nuclear strands, irrespective of the positions or order of the query sequences in the organelle genome (Supplementary Table 9).

Candidate LGTs in eukaryote genomes. The number of cLGTs specific to each eukaryote genome was estimated by BLAST[52] version 2.2.26 searches using all prokaryotic protein sequences and the eukaryotic proteins that were not clustered with any protein from another eukaryote (that is, those found neither in ESCs nor in EPCs). The number of protein sequences with at least one prokaryote hit

(E value $\leq 1 \times 10^{-5}$, identity $\geq 95\%$) was reported for each eukaryotic genome (Supplementary Table 9).

Eukaryote non-monophyly in phylogenetic trees. In this study we detected 1,933 EPCs that recovered eukaryotic monophyly in maximum likelihood trees in addition to 329 EPCs that did not reject eukaryote monophyly in AUTs (Extended Data Fig. 1b). The remaining 323 EPCs produced maximum likelihood trees in which the eukaryotic sequences neither formed a monophyletic group nor passed the AUT (Extended Data Fig. 1b). It is possible that these 323 trees represent LGTs, but it is also possible that factors pertaining to the inference of phylogenetic trees are responsible for the failure of the eukaryotic sequences to form a monophyletic group. At least three well-known classes of factor can cause a proportion of eukaryote genes to branch in a non-monophyletic manner in molecular phylogenies: biological causes (for example, host and endosymbiont copies of a given gene persist), contamination in genome sequences, and limitations of phylogenetic methods.

First, among the 323 non-monophyly cases, biological causes constitute a significant class. It is uncontested that, during eukaryotic evolution, endosymbiosis brought together at least three different prokaryotic partners, which served as sources of nuclear genes: cyanobacteria, alphaproteobacteria, and archaea (Fig. 2). For essential cellular functions that were common to both endosymbiont and host such as ribosome biogenesis, amino-acid biosynthesis, nucleotide biosynthesis, cofactor biosynthesis, or carbohydrate metabolism, endosymbiosis brings together divergent but often homologous gene copies within the same cell. This occurs both at the origin of mitochondria and at the origin of plastids (including secondary symbiosis). The phenomenon, called functional redundancy through endosymbiosis[87], is reasonably well known. It often happens that both a host copy and an endosymbiont copy persist in a given eukaryotic lineage, ribosomal proteins being one example[88], chloroplast–cytosol isoenzymes being another[87]. Such homologous gene copies, sequence conservation permitting, can come to reside within the same EPC. Within the 323 non-monophyly cases (Supplementary Table 6), 218 genes (67%) are involved in such essential function: 38 genes (trees) are involved in ribosome biogenesis (including 19 ribosomal proteins), 55 in amino-acid metabolism, 27 in carbohydrate metabolism, 23 in nucleotide metabolism, 16 in cofactor metabolism, 33 in energy conservation, 11 in lipid metabolism, and 13 in post-translational modification. In cases of symbiotic redundancy, if copies from more than one symbiotic partner persist in any eukaryotic lineage sampled, eukaryotic sequences will form two or three distinct clades in the trees, if, that is, that phylogeny is reconstructed accurately in that regard. Before it was known how widespread LGT among prokaryotes is, there was an expectation that genes affected by symbiotic redundancy should branch with cyanobacterial and alphaproteobacterial homologues[87], but that expectation turned out to be too optimistic (Fig. 2) and has been revised[79]. Many of the 323 non-monophyly cases will ultimately be attributable to symbiotic redundancy, but it is not our aim to present that interpretation here. In addition to patterns suggesting LGT to eukaryotes, eukaryote non-monophyly patterns suggesting LGT from eukaryotes to prokaryotes were also observed. Many prokaryotes can take up foreign DNA present in the environment[1,3,89]. Among the 323 cases of non-monophyly, 21 trees show prokaryotic sequences nested within a eukaryote clade (Supplementary Table 6).

Second, bacterial contaminations during genome sequencing will generate non-monophyletic trees for eukaryotes (prokaryotic sequences with eukaryotic taxon labels). We took the data from the genomes as it was, without cleaning or purging for possible contaminations, which would have biased our results towards eukaryote monophyly in trees. Probable cases of contaminating DNA could be found in the eukaryote genome sequence data used in this study. In 78 trees, eukaryotes were non-monophyletic owing to the presence of only one or two eukaryotic outlier organisms. A notable source of outliers is the genome sequence of the sea anemone Nematostella[90], which was shown to contain sequences from Proteobacteria and Bacteroidetes[91]. In eukaryote non-monophyly EPC trees, putative contaminations in Nematostella were often found as the single outlier (7 out of 52, 13%; Supplementary Table 6) or together with an additional outlier (6 out of 28, 21%; Supplementary Table 6), frequently with either Proteobacteria (for example, E6978_B51) or Bacteroidetes (for example, E3129_B78) taxa in its sister group. Further evidence for contaminating DNA in the Nematostella genome comes from the observation that over half of the cLGTs in the 55 genomes stem from the Nematostella sequences (Supplementary Table 9). Another source of putative prokaryotic contaminations is the sponge Amphimedon[92], an organism known to have dense communities of symbiotic prokaryotes, which could be sources of bacterial contaminants as a result of sequence misassembly[93]. In 9 out of 52 (17%) eukaryote non-monophyly EPC trees with a single eukaryotic outlier organism, and in 9 out of 28 (32%) trees with two eukaryotic outlier organisms, Amphimedon (Supplementary Table 6) was an outlier. Single Amphimedon outliers in the eukaryote non-monophyly EPC trees

tend to be nested within a clade of gammaproteobacterial sequences as a long-branch (for example, E841_B491, E869_B486, E3655_B52). This is suggestive of the fast-evolving characteristic of symbiotic bacteria[94] and explains why, in contrast to *Nematostella*, the cLGT detection approach (BLAST local identity ≥95%) revealed no cLGT in *Amphimedon* (Supplementary Table 9), despite these putative contaminating bacterial sequences revealed by the trees. In addition, 32 eukaryote non-monophyly trees contain only two eukaryotic organisms, with *Amphimedon* and/or *Nematostella* accounting for 50% of those occurrences (Supplementary Table 6). Although putative contaminations are especially abundant in aquatic organisms or organisms with symbiotic prokaryotes, such as the known case of *Hydra* endosymbiotic bacterial contaminants[95], they can also be found in multicellular land organisms, such as mammals[96] or plants[97]. Contaminations need not stem from the DNA sample sequenced, but can also be introduced from vectors during the sequencing process[97]. The same putative contamination can even be present in genome sequences of different eukaryotes through the use of similar sequencing procedures. An example might be the EPC E14272_B12261, where a transposase gene only present in *Oryza* and *Trypanosoma* (both sequenced using the bacterial artificial chromosome) is 100% identical to the *E. coli* homologue. We used the genome data without purging for possible contaminations, which are, however, present in the data.

Third, factors affecting phylogeny can generate eukaryote non-monophyly in trees. Phylogenetic algorithms strive to find the best tree under a given evolutionary model[22,23,34]. If the model is misspecified, the best tree by a likelihood criterion need not be the true tree[25]. In eukaryote evolution, the duplication of genes and whole genomes is a very frequent phenomenon[98]. In duplicated families, functional constraints can change across sequence positions and across subfamilies, leading to covarion/covariotide phenomena (heterogeneity of the substitution process across sites and across the tree), which can generate phylogenetic artefacts, especially when gene duplicates are present[34,99,100]. We counted the number of EPCs in which any eukaryote was represented with more than one sequence. Among the 323 eukaryote non-monophyletic clusters that failed the AUT, such EPCs are overrepresented in comparison with monophyletic clusters (χ^2 goodness-of-fit test, $P = 6.06 \times 10^{-11}$; Supplementary Table 6). A significant, although much higher, P value was obtained for non-monophyletic clusters that passed the AUT ($P = 3.47 \times 10^{-4}$; Supplementary Table 6). Sampling is also an issue for phylogenetic analyses. We found 23 cases where cyanobacterial sequences were nested within the photosynthetic eukaryotes and their relatives (7 additional cases in which an outlier, possible sequencing contamination, appeared in the tree; Supplementary Table 6). Tree E1689_B206_A295 for example, contains 1,746 sequences and fails the AUT for eukaryote monophyly; however, adding merely ten new top BLAST[52] prokaryote hits from the most recent NR database[101] using the *Arabidopsis* sequence as the query (as of 17 April 2015), produces a highest likelihood tree with Archaeplastida monophyly (Extended Data Fig. 3). That taxon sampling affects phylogeny is well-known[102]; it affects all analyses, not just the present one. Another factor is clustering. Clustering and alignment can introduce phylogenetic biases; larger clusters produce eukaryote non-monophyly significantly more often than smaller clusters ($P = 1.45 \times 10^{-61}$) as do trees generated from the least reliable alignments ($P = 2.04 \times 10^{-10}$; Extended Data Fig. 2). The two-step clustering procedure used in this study avoids combining sequences into families that are too large and complex in terms of shared protein domains: the joining of a cluster for protein A to a cluster for protein B via a single AB fusion protein generates extremely large families, sometimes called giant connected components[103]. However, the universal identity threshold across all clusters could result in over-clustering in some cases: grouping of distinct prokaryotic families, each with eukaryotic homologues, into a single cluster with two eukaryotic branches, each monophyletic, but generating eukaryote non-monophyly for the cluster.

For 134 trees, there was no obvious contamination problem or case of cyanobacteria and plants interleaving. These 134 cases were therefore classified as putative LGT (Supplementary Table 6). But when the 134 cases were compared with the eukaryote monophyletic EPCs, we found significantly more trees than expected with any eukaryote having more than one gene copy (duplicates) ($P = 1.72 \times 10^{-13}$; Supplementary Table 6); in the remaining 189 cases the P value increased to 4×10^{-3}. The presence of an additional, divergently branching copy can result from functional redundancy through endosymbiosis[87] and differential loss, through heterogeneity of the substitution process across sites and across the tree[34,99,100], or through lineage-specific LGT. Of course, many of the trees in question might be affected by more than one of these factors. If LGT is the cause of these 323 cases, which for this paper we conservatively assume, then the eukaryotes in question are still not expanding their gene repertoire, they are merely reacquiring fresh copies of genes already present in the eukaryotic lineage. The details of these 323 trees are in Supplementary Table 6; the trees themselves are in Supplementary Table 7.

Estimating the relative contributions of the host, mitochondria, and plastids to the gene repertoire of present-day eukaryotes. The proportion of genes contributed by the archaeal host is calculated as the proportion of eukaryote monophyly EPC trees where archaea are found in the sister group, including 314 with pure archaeal sister groups and 33 with both archaea and bacteria in the sister group (Extended Data Fig. 5): 347/2,585 = 13.42%. The contribution from the plastid ancestor is calculated by regarding all clusters in the ABC block (Fig. 1) as genes of plastid origin other than those (83) where eukaryotes are monophyletic with archaea in the sister group: (1,060 − 83)/2,585 = 37.79%. The mitochondrion-derived genes are all the other genes: 100% − 13.42% − 37.79% = 48.79%.

Note that the number for the host contribution is probably an underestimate, as only EPCs with a monophyletic eukaryotic clade in the maximum likelihood tree were counted. For genes of plastid origin, it might be a slight overestimate, since there would also be genes of plastid–host origin that are now specific to Archaeplastida/SAR/Hacrobia and found in the ABC block as the result of differential loss. Another complication is that there can be clusters with genes from more than one source (see above), so there can be, for example, E block clusters of partial plastid and partial mitochondrial origin.

51. Nelson-Sathi, S. *et al.* Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature* **517,** 77–80 (2015).
52. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
53. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278,** 631–637 (1997).
54. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276–277 (2000).
55. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).
56. Apic, G., Gough, J. & Teichmann, S. A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310,** 311–325 (2001).
57. Powell, S. *et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* **42,** D231–D239 (2014).
58. Tatusov, R. *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinform.* **4,** 41 (2003).
59. Yoon, H. S., Muller, K. M., Sheath, R. G., Ott, F. D. & Bhattacharya, D. Defining the major lineages of red algae (Rhodophyta). *J. Phycol.* **42,** 482–492 (2006).
60. James, T. Y. *et al.* Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* **443,** 818–822 (2006).
61. Okamoto, N., Chantangsi, C., Horak, A., Leander, B. S. & Keeling, P. J. Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLoS ONE* **4,** e7080 (2009).
62. Hampl, V. *et al.* Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc. Natl Acad. Sci. USA* **106,** 3859–3864 (2009).
63. Janouškovec, J., Horák, A., Oborník, M., Lukeš, J. & Keeling, P. J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl Acad. Sci. USA* **107,** 10949–10954 (2010).
64. Lahr, D. J. G., Grant, J., Nguyen, T., Lin, J. H. & Katz, L. A. Comprehensive phylogenetic reconstruction of Amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. *PLoS ONE* **6,** e22780 (2011).
65. Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59,** 429–493 (2012).
66. Leliaert, F. *et al.* Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* **31,** 1–46 (2012).
67. Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64,** 583–607 (2013).
68. Jackson, C. J. & Reyes-Prieto, A. The mitochondrial genomes of the glaucophytes *Gloeochaete wittrockiana* and *Cyanoptyche gloeocystis*: multilocus phylogenetics suggests a monophyletic Archaeplastida. *Genome Biol. Evol.* **6,** 2774–2785 (2014).
69. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
70. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
71. Yutin, N. & Galperin, M. Y. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* **15,** 2631–2641 (2013).
72. Landan, G. & Graur, D. Heads or tails: a simple reliability check for multiple sequence alignments. *Mol. Biol. Evol.* **24,** 1380–1383 (2007).
73. Landan, G. & Graur, D. Local reliability measures from sets of co-optimal multiple sequence alignments. *Pacif. Symp. Biocomput.* **13,** 15–24 (2008).
74. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17,** 1246–1247 (2001).
75. Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53,** 131–147 (1981).
76. Felsenstein, J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266,** 418–427 (1996).

77. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
78. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
79. Ku, C. *et al.* Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. *Proc. Natl. Acad. Sci. USA* (2015).
80. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57,** 289–300 (1995).
81. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29,** 1165–1188 (2001).
82. Zar, J. H. *Biostatistical Analysis* Ch. 22 (Pearson, 2014).
83. Dagan, T. & Martin, W. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl Acad. Sci. USA* **104,** 870–875 (2007).
84. Petitjean, C., Deschamps, P., Lopez-Garcia, P. & Moreira, D. Rooting the domain Archaea by phylogenomic analysis supports the foundation of the new kingdom Proteoarchaeota. *Genome Biol. Evol.* **7,** 191–204 (2015).
85. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform.* **10,** 421 (2009).
86. Hazkani-Covo, E. & Graur, D. A comparative analysis of numt evolution in human and chimpanzee. *Mol. Biol. Evol.* **24,** 13–18 (2007).
87. Martin, W. & Schnarrenberger, C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32,** 1–18 (1997).
88. Maier, U. G. *et al.* Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. *Genome Biol. Evol.* **5,** 2318–2329 (2013).
89. de Vries, J. & Wackernagel, W. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc. Natl Acad. Sci. USA* **99,** 2094–2099 (2002).
90. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
91. Artamonova, I. I. & Mushegian, A. R. Genome sequence analysis indicates that the model eukaryote *Nematostella vectensis* harbors bacterial consorts. *Appl. Environ. Microbiol.* **79,** 6868–6873 (2013).
92. Srivastava, M. *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466,** 720–726 (2010).
93. Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nature Rev. Microbiol.* **10,** 641–654 (2012).
94. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nature Rev. Microbiol.* **10,** 13–26 (2012).
95. Wenger, Y. & Galliot, B. RNAseq versus genome-predicted transcriptomes: a large population of novel transcripts identified in an Illumina-454 *Hydra* transcriptome. *BMC Genom.* **14,** 204 (2013).
96. Langdon, W. B. Mycoplasma contamination in the 1000 Genomes Project. *BioData Min.* 7, 3 (2014).
97. Lang, D., Zimmer, A. D., Rensing, S. A. & Reski, R. Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends Plant Sci.* **13,** 542–549 (2008).
98. Maere, S. *et al.* Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA* **102,** 5454–5459 (2005).
99. Lockhart, P. J., Larkum, A. W. D., Steel, M. A., Waddell, P. J. & Penny, D. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. USA* **93,** 1930–1934 (1996).
100. Lockhart, P. J. *et al.* How molecules evolve in eubacteria. *Mol. Biol. Evol.* **17,** 835–838 (2000).
101. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33,** D501–D504 (2005).
102. Zwickl, D. J. & Hillis, D. M. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* **51,** 588–598 (2002).
103. Alvarez-Ponce, D., Lopez, P., Bapteste, E. & McInerney, J. O. Gene similarity networks provide tools for understanding eukaryote origins and evolution. *Proc. Natl. Acad. Sci. USA* **110,** E1594–E1603 (2013).
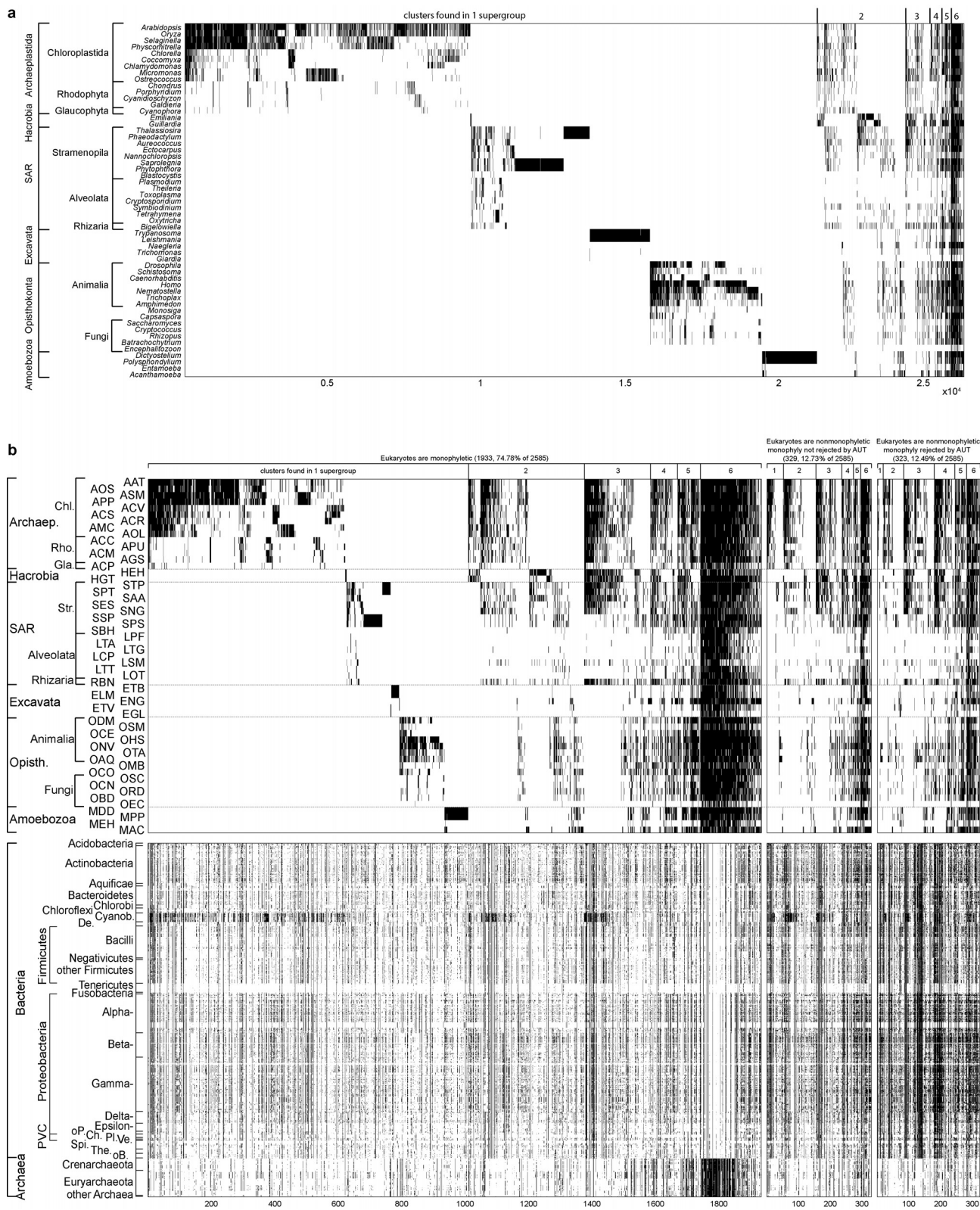
Extended Data Figure 1 | Additional gene distribution patterns.
a, Distribution of ESCs. Each black tick indicates the presence of a cluster in a taxon. The 26,117 ESCs (x axis) from 55 eukaryotic genomes (Supplementary Table 1) are sorted according to their distribution across the six eukaryotic supergroups. **b**, Distribution of taxa in EPCs and monophyly of eukaryotes. Each black tick indicates the presence of a cluster in a taxon. The 2,585 EPCs (x axis) are separated into three sets according to the monophyly of eukaryotes and the results of the AUT and, within each set, are ordered according to

their distribution across the six eukaryotic supergroups. Clusters where eukaryotes were resolved as non-monophyletic in the maximum likelihood tree tend to occur more frequently in bacterial taxa. Archaep., Archaeplastida; Opisth., Opisthokonta; Chl., Chloroplastida; Rho., Rhodophyta; Gla., Glaucophyta; Str., Stramenopila; De., Deinococcus-Thermus; oP., other Proteobacteria; Ch., Chlamydiae; Pl., Planctomycetes; Ve., Verrucomicrobia; Spi., Spirochaetae; The., Thermotogae; oB., other Bacteria. For abbreviations of eukaryotes, see Supplementary Table 1.

a

b

c

d
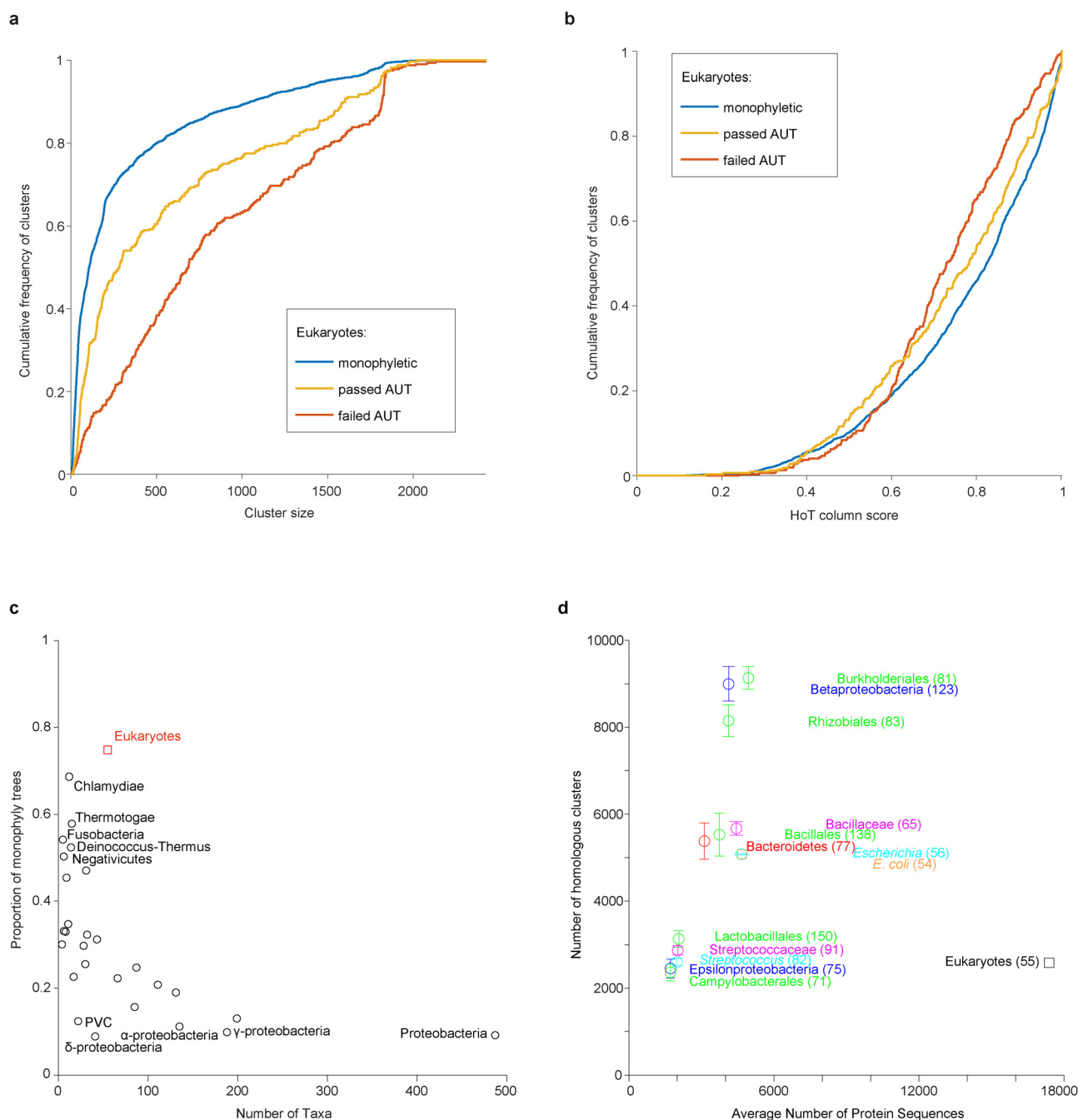
Extended Data Figure 2 | Clustering, monophyly, and gene sharing.
a, **b**, Monophyly of eukaryotes in maximum likelihood trees, cluster size, and alignment quality. Cumulative frequency of clusters with different cluster size (**a**) or different HoT[72] column scores (**b**) is plotted for three sets of EPCs that differ in terms of the monophyly of eukaryotes in the maximum likelihood trees (monophyletic: resolved as monophyletic in the original tree; passed AUT: resolved as non-monophyletic in the original tree, but at least one alternative tree with eukaryote monophyly (see Methods) was as likely at $P = 0.05$ in an AUT; failed AUT: alternative trees were not as likely as the original tree where eukaryotes were resolved as non-monophyletic). One-sided Kolmogorov–Smirnov two-sample goodness-of-fit test (cluster size/HoT column scores): monophyletic versus passed AUT, $1.04 \times 10^{-13}/7.9 \times 10^{-3}$; monophyletic versus failed AUT, $1.45 \times 10^{-61}/2.04 \times 10^{-10}$; passed AUT versus failed AUT, $3.40 \times 10^{-13}/4.00 \times 10^{-3}$. **c**, **d**, Prokaryotic monophyly and gene sharing. **c**, Proportion of trees showing monophyly for taxonomic group.

Prokaryotic phyla and classes (Supplementary Tables 3 and 4) that are monophyletic in the reference trees and that have at least five taxa (genomes in archaea or species in bacteria) are plotted according to the number of taxa and the proportion of EPC trees with at least two sequences from a prokaryotic group where it forms a monophyletic group. The proportion of eukaryote monophyly trees is higher than that of any prokaryotic group, including those with many fewer taxa. **d**, Gene sharing between a prokaryotic group and other prokaryotes. Using the same procedure for the generation of EPCs, 55 genomes were randomly sampled from a group of bacteria and the number of clusters (EPCs) they shared with prokaryotes not from this group was counted. The average number of shared clusters was mapped for each taxonomic group with 55–150 genomes (error bar, s.d.; number of genomes in parentheses). For *E. coli* and the eukaryotes (shown for comparison), there was only one sample. Colour coding for taxonomic levels: red, phylum; blue, class; green, order; magenta, family; cyan, genus; orange, species.
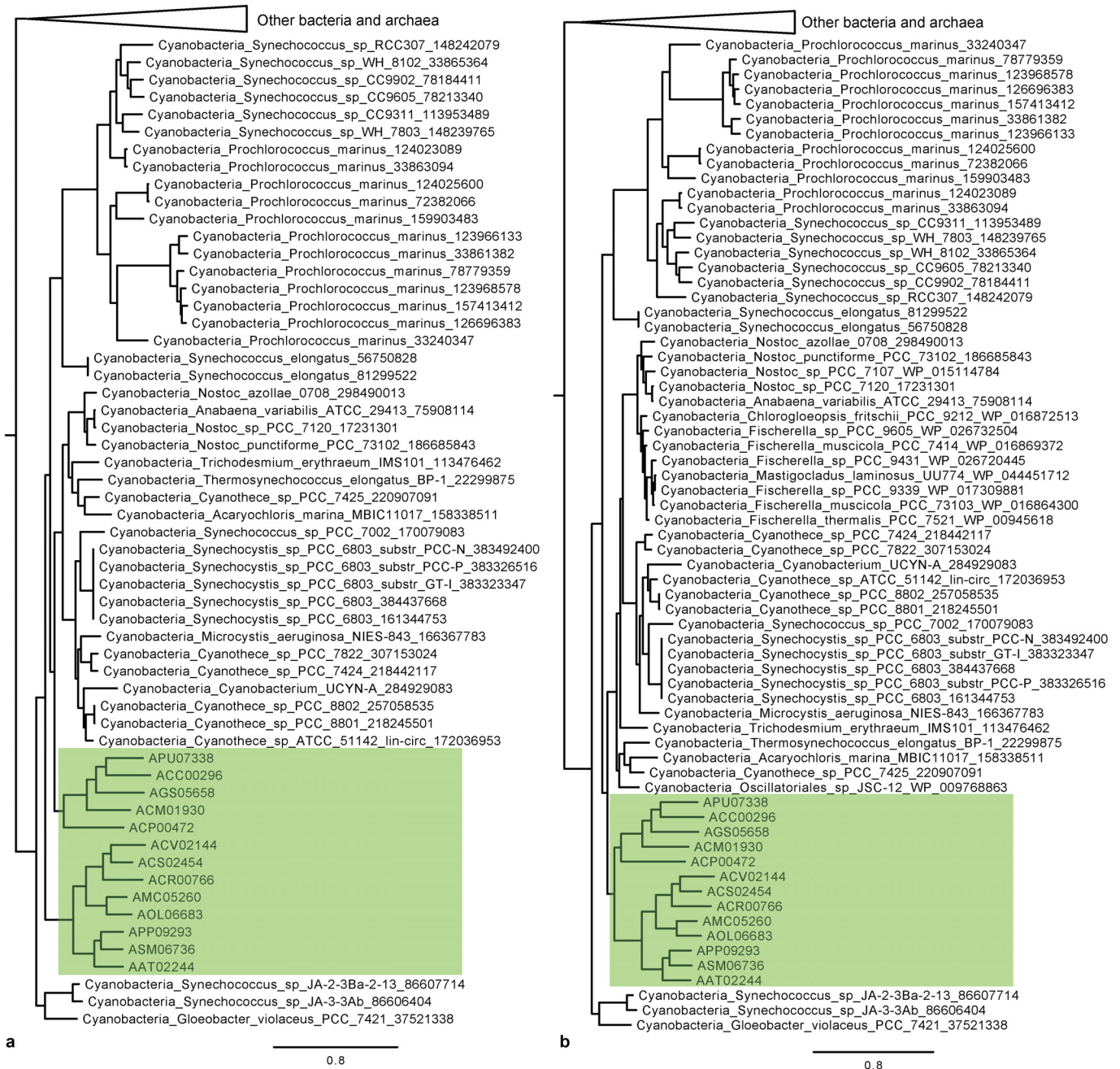
Extended Data Figure 3 | Effect of taxon sampling on eukaryote monophyly in phylogenetic trees. After ten sequences (bold) were added to the original data set (EPC E1689_B206_A295), the relationships among Archaeplastida taxa (highlighted in green) changed from non-monophyly (**a**) to monophyly (**b**).

Abbreviations are shown for eukaryotic sequences (Supplementary Table 2) and NCBI GI numbers for cyanobacterial sequences (Supplementary Table 3; RefSeq accessions are shown for the added sequences).
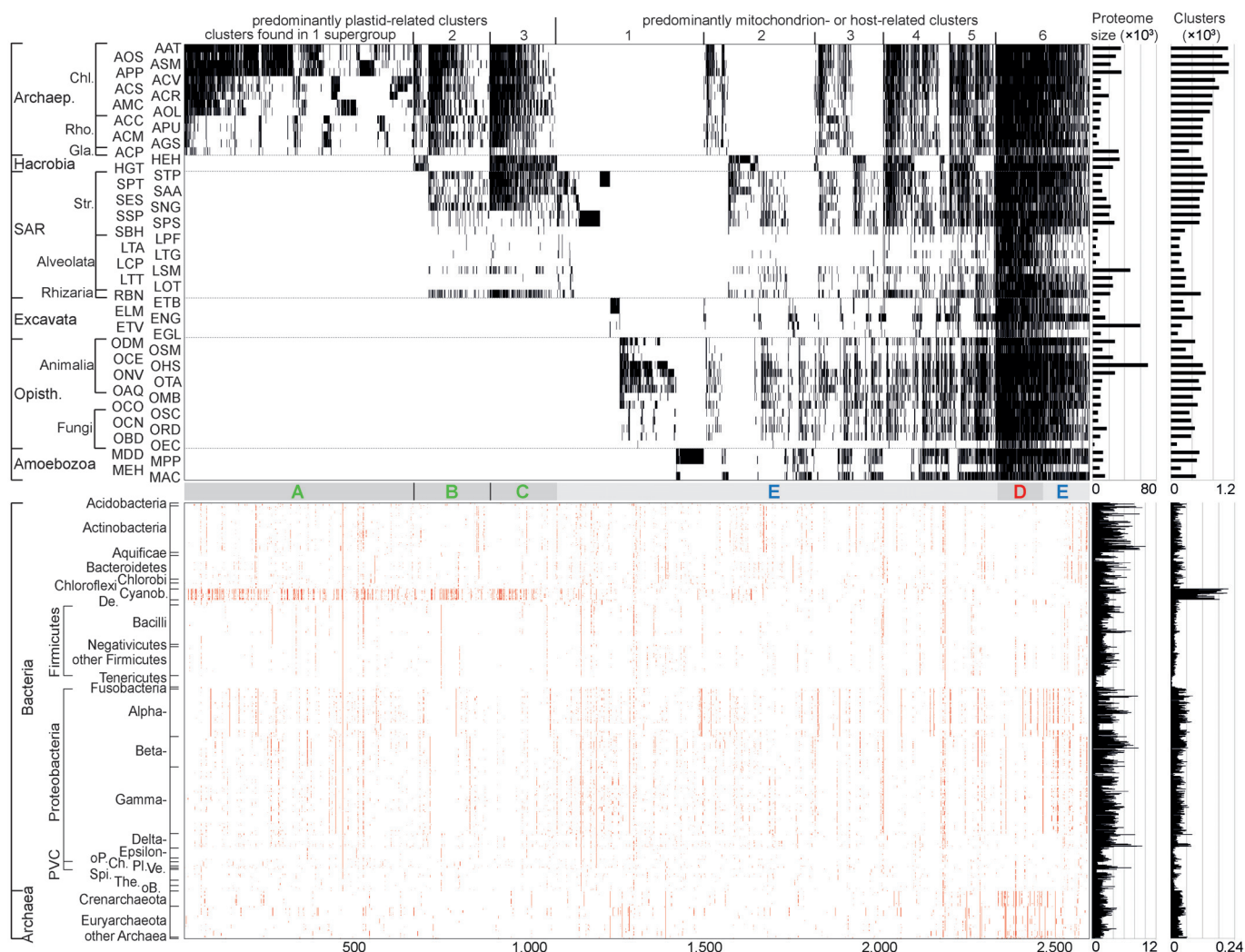
Extended Data Figure 4 | Distribution of prokaryotic taxa in the sister group to eukaryotes, with EPCs sorted by eukaryotic supergroups. Top: each black tick indicates the presence of a eukaryote taxon in one of the 2,585 EPCs. Bottom: each red tick indicates the presence of a prokaryote taxon in the sister group to eukaryotes in one of the 1,933 EPC maximum likelihood trees where eukaryotes were resolved to be monophyletic. The 2,585 EPCs, proteome size, and cluster size are as in Fig. 1. The number of EPCs present and the frequency of occurrence in the sister group to eukaryotes ('clusters') are shown for eukaryotes and prokaryotes, respectively. Archaep., Archaeplastida; Opisth., Opisthokonta; Chl., Chloroplastida; Rho., Rhodophyta; Gla., Glaucophyta; Str., Stramenopila; De., Deinococcus-Thermus; oP., other Proteobacteria; Ch., Chlamydiae; Pl., Planctomycetes; Ve., Verrucomicrobia; Spi., Spirochaetae; The., Thermotogae; oB., other Bacteria. For abbreviations of eukaryotes, see Supplementary Table 1.
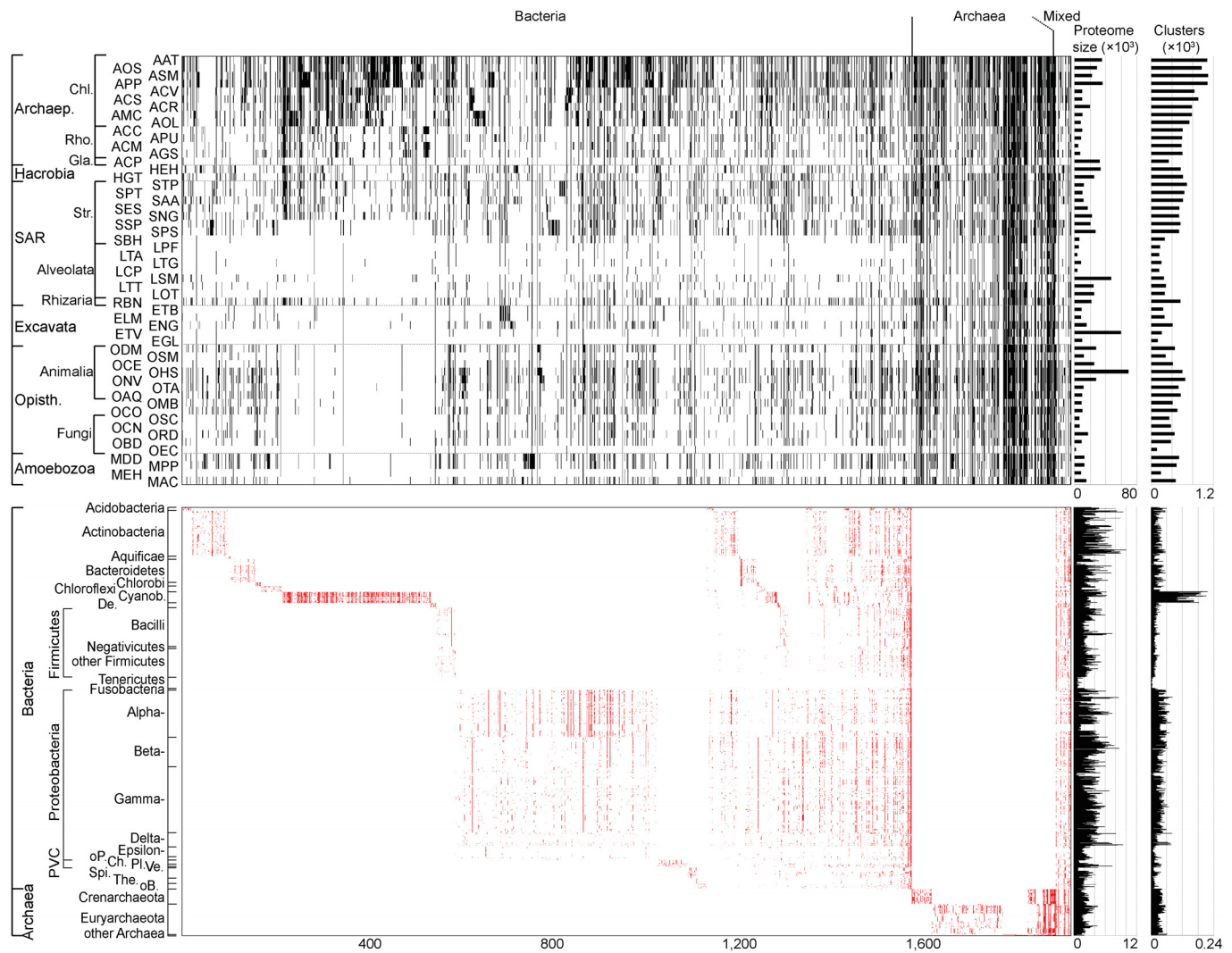
Extended Data Figure 5 | Distribution of prokaryotic taxa in the sister group to eukaryotes, with EPCs sorted by prokaryotic groups. Top: each black tick indicates the presence of a eukaryote taxon in one of the 1,933 EPC maximum likelihood trees where eukaryotes were resolved to be monophyletic. Bottom: each red tick indicates the presence of a prokaryote taxon in the sister group to eukaryotes in one of those 1,933 EPC trees. The EPCs (x axis) are ordered according to the taxonomic groups to which the prokaryotes in the sister group to eukaryotes belong (separated into three blocks where only bacteria (1,586 EPCs), only archaea (314 EPCs), or both bacteria and archaea (33 EPCs) are found in the sister group). There are 16 bacterial groups

(including 'other Bacteria'; Firmicutes, Proteobacteria, and the PVC superphylum (Planctomycetes, Verrucomicrobia, and Chlamydiae) are regarded as single groups) and five archaeal groups (the five phyla). The number of EPCs present and the frequency of occurrence in the sister group to eukaryotes are shown for eukaryotes and prokaryotes, respectively. Archaep., Archaeplastida; Opisth., Opisthokonta; Chl., Chloroplastida; Rho., Rhodophyta; Gla., Glaucophyta; Str., Stramenopila; De., Deinococcus-Thermus; oP., other Proteobacteria; Ch., Chlamydiae; Pl., Planctomycetes; Ve., Verrucomicrobia; Spi., Spirochaetae; The., Thermotogae; oB., other Bacteria. For abbreviations of eukaryotes, see Supplementary Table 1.
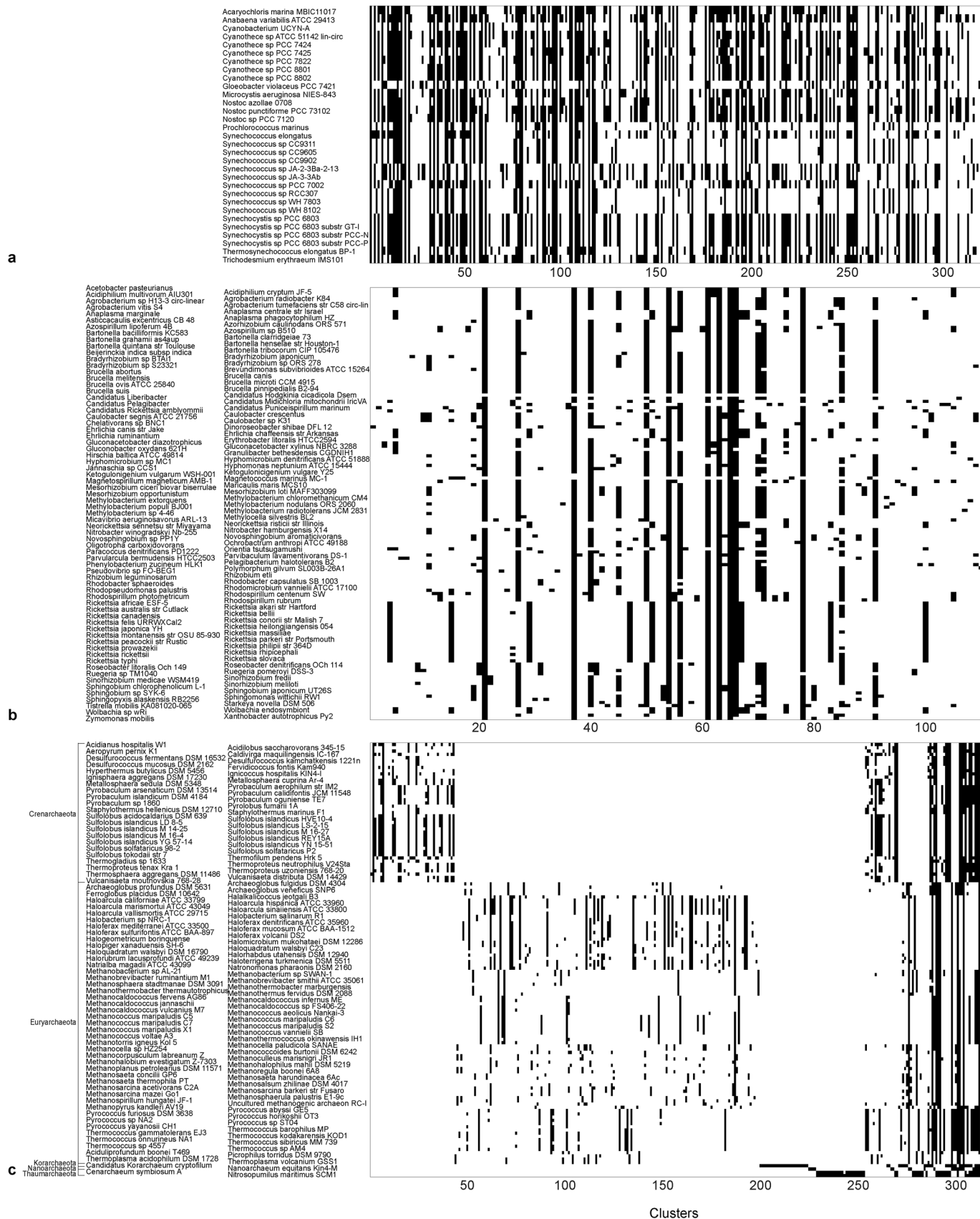
Extended Data Figure 6 | Distribution of taxa in the sister groups consisting purely of cyanobacteria, alphaproteobacteria, or archaea. Each black tick indicates the presence of a prokaryotic taxon in the sister group to eukaryotes in an EPC tree. **a–c**, Distributions of taxa in all pure-cyanobacterial (**a**), pure-alphaproteobacterial (**b**), and pure-archaeal (**c**) sister groups. The clusters are ordered alphanumerically according to the eukaryotic cluster numbers (Supplementary Table 5), whereas for archaea (**c**) the taxa are further sorted by the five archaeal phyla.
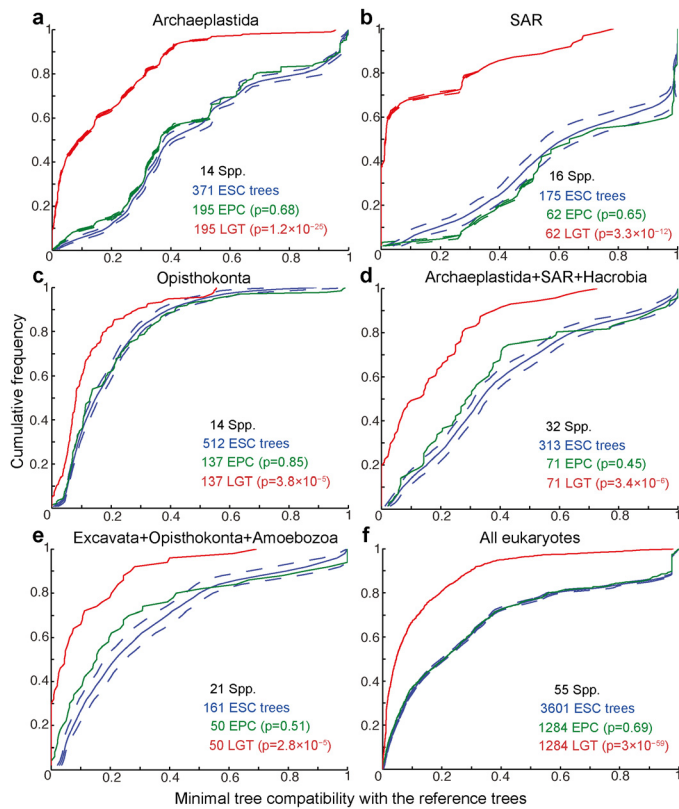
Extended Data Figure 7 | Comparison of sets of trees for single-copy genes in eukaryotic groups, with more inclusive criteria. a–f, Cumulative distribution functions (*y* axis) for scores of minimal tree compatibility with the vertical reference data set (*x* axis). Values are number of species, sample sizes, and *P* values of the two-tailed Kolmogorov–Smirnov two-sample goodness-of-fit test in the comparison of the ESC (blue) data sets against the EPC (green) data set and a synthetic data set simulating one LGT (red). Dashed lines delineate the range of distributions in 100 replicates of random down-sampling. The criteria for tree inclusion were less stringent than those for Fig. 3 (see Methods).
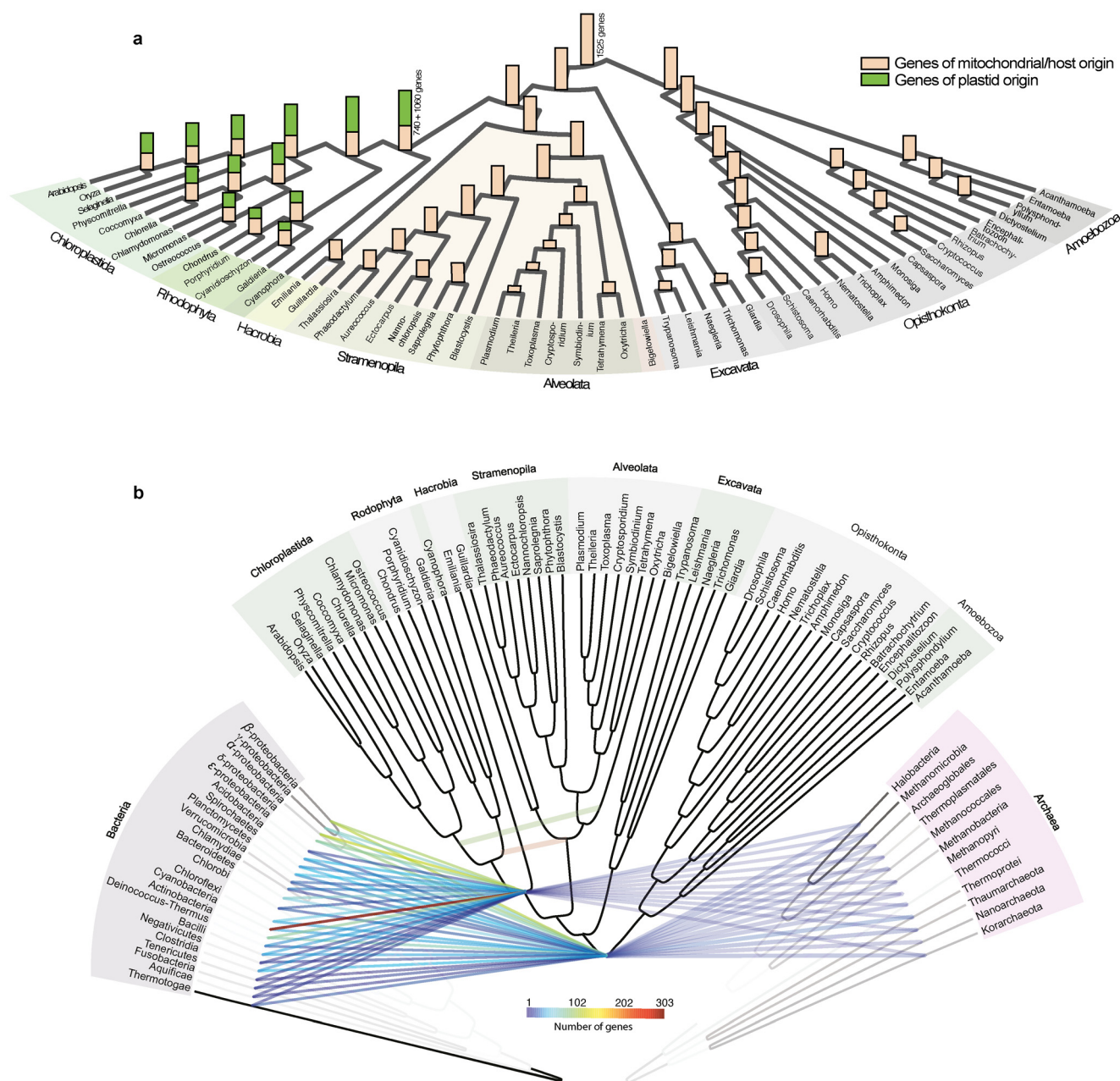
Extended Data Figure 8 | Overview of eukaryote gene content evolution.
a, Eukaryotic evolution by gene loss. Genome sizes (number of EPCs present) were mapped onto the eukaryotic reference tree. Ancestral genome size in each eukaryotic ancestral node was calculated using a loss-only model, with all EPCs in blocks A–C and those in blocks D and E (Fig. 1) entering the eukaryotic lineage via the plastid ancestor (green) or the eukaryote ancestor (wheat colour). Plastid-derived genes are not shown for the ancestral nodes within SAR and Hacrobia, because of current debates about the number and nature of secondary symbioses, but are indicated by the greenish shading. **b**, Endosymbiotic gene transfer network. The network connecting apparent gene donors to the common ancestor of eukaryotes and Archaeplastida is mapped onto the reference phylogeny (vertical edges) of bacteria (left),

eukaryotes (middle), and archaea (right). Grey shading (white to black) in the prokaryote reference trees (70 for archaea and 32 for bacteria) indicates how often a branch associated with a particular node was recovered within the trees of individual genes that were concatenated for inferring the reference topology. Lateral edges indicate gene influx at the origin of eukaryotes and at the origin of plastids. Edge colour corresponds to the frequencies with which a prokaryotic group appears in the sister group to eukaryotes. The archaeal reference tree was rooted between euryarchaeotes and other taxa, and the bacterial tree with Thermotogae. Secondary endosymbiotic transfers are indicated in light green and red. That members of both the Crenarchaeota and the Euryarcheaota are implicated as host relatives is probably because of the small archaeon sample[34–36].
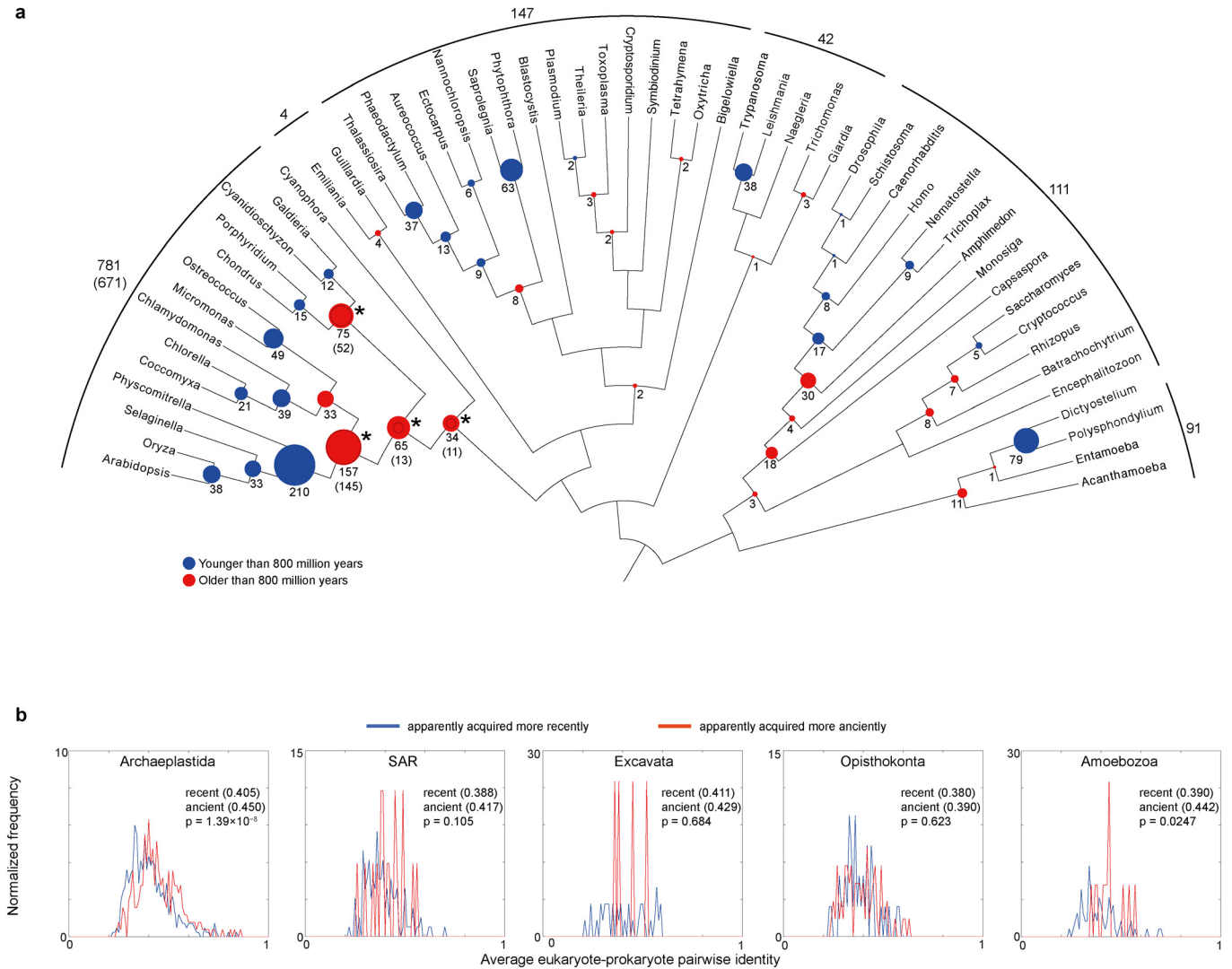
Extended Data Figure 9 | Apparent gene transfers and eukaryote–prokaryote sequence identities. a, Patterns suggestive of LGT from prokaryotes inferred from EPC trees. All EPC trees were searched for phylogenetic patterns suggestive of gene acquisitions by the common ancestor of each eukaryote lineage within the six supergroups (see Methods). The size of each circle is proportional to the number of such putative acquisitions, with the total number of putative acquisitions shown for each supergroup. The colour shows the age of nodes according to a eukaryotic time tree (blue, younger than 800 million years; red, older than 800 million years). For the four lineages with an asterisk, phylogenetic patterns where SAR/Hacrobia are nested within a clade formed by Archaeplastida were also counted as putative acquisitions to take into account secondary plastid endosymbioses. The numbers of acquisitions without such patterns are indicated in parentheses (and shown as inner circles). **b**, Eukaryote–prokaryote sequence identities for genes apparently acquired more recently and more anciently in eukaryotes (**a**). The mean of the average pairwise identities is shown in parentheses. At $P = 0.05$, a two-sided Wilcoxon rank-sum test either did not reject the null hypotheses that the two sets of genes are not different or suggested the tip-specific eukaryotic genes are less similar to their prokaryotic homologues.

Extended Data Figure 10 | Distribution of ESCs and EPCs across eukaryotes under different criteria. Different thresholds were applied to find eukaryote clusters with prokaryote homologues, including BLAST local identity for each eukaryote–prokaryote hit (30% or 20%) and levels of best-hit correspondence (10–50%) for identifying reciprocal pairs of eukaryote and prokaryote clusters. Distributions of ESCs and EPCs are drawn as in Extended Data Fig. 1a and Fig. 1, respectively.

ARTICLE

Structure and mechanism of an active lipid–linked oligosaccharide flippase

Camilo Perez[1], Sabina Gerber[1]†, Jérémy Boilevin[2], Monika Bucher[1], Tamis Darbre[2], Markus Aebi[3], Jean-Louis Reymond[2] & Kaspar P. Locher[1]

The flipping of membrane-embedded lipids containing large, polar head groups is slow and energetically unfavourable, and is therefore catalysed by flippases, the mechanisms of which are unknown. A prominent example of a flipping reaction is the translocation of lipid-linked oligosaccharides that serve as donors in N-linked protein glycosylation. In _Campylobacter jejuni_, this process is catalysed by the ABC transporter PglK. Here we present a mechanism of PglK-catalysed lipid-linked oligosaccharide flipping based on crystal structures in distinct states, a newly devised _in vitro_ flipping assay, and _in vivo_ studies. PglK can adopt inward- and outward-facing conformations _in vitro_, but only outward-facing states are required for flipping. While the pyrophosphate-oligosaccharide head group of lipid-linked oligosaccharides enters the translocation cavity and interacts with positively charged side chains, the lipidic polyprenyl tail binds and activates the transporter but remains exposed to the lipid bilayer during the reaction. The proposed mechanism is distinct from the classical alternating-access model applied to other transporters.

The translocation of lipids across the bilayer, termed flipping, is not only essential for maintaining lipid asymmetry in membranes, but also underpins various processes including signalling, vesicle formation in secretory and endocytic pathways, the regulation of membrane protein activity, and asparagine-linked protein glycosylation[1–5]. In animals, _trans_-bilayer lipid asymmetry affects processes such as blood coagulation, macrophage recognition[6] and apoptosis[7]. In bacteria, precursors of cell wall components are coupled to lipid carriers and exported by flippases, which is essential for cell wall assembly[8–10]. Flipping reactions are catalysed by passive or active transporter proteins[11]. These include energy-independent scramblases that randomize the orientation of lipids across the bilayer[12,13], proton- or sodium-driven secondary antiporters[14], and ATP-driven, vectorial transporters that actively translocate specific lipids. The latter include P4-ATPases[15] and ATP-binding cassette (ABC) transporters[11,16].

Despite extensive research, remarkably little is known about the structure and mechanism of active, flippase-catalysed lipid translocation. Until now, crystal structures have only been reported for the MsbA protein[16,17], the bacterial flippase of lipid A-core, but no mechanism could be deduced from the structural data. In addition to the scarcity of structures, there are few reliable _in vitro_ assays that allow the study of native substrate flipping.

A prominent substrate of flippases in eukaryotes is the lipid-linked-oligosaccharide (LLO) Man$_5$GlcNAc$_2$-PP-dolichol, which is translocated from the cytosolic to the luminal side of the endoplasmic reticulum membrane, where the oligosaccharide is transferred to asparagine residues of nascent polypeptide chains in protein N-glycosylation[1,2]. In the homologous bacterial N-glycosylation pathway, a chemically similar LLO (GlcGalNAc$_5$Bac-PP-undecaprenyl; Extended Data Fig. 1) is flipped in the human pathogen _C. jejuni_[18,19]. The responsible flippase, PglK, is a homodimeric ABC transporter and an essential component of the bacterial protein N-glycosylation machinery[20]. The translocated LLO is handed over to the oligosaccharyltransferase PglB, which catalyses oligosaccharide transfer to acceptor proteins[21]. Undecaprenyl-pyrophosophate is subsequently

hydrolysed to the monophosphate by undecaprenyl-pyrophosphatase[22] (Extended Data Fig. 2).

To understand the mechanism of ATP-driven LLO flipping, we have developed an _in vitro_ assay and determined the crystal structures of _C. jejuni_ PglK in distinct states. Our results suggest a mechanism in which only the outward-facing conformations are relevant for flipping, thereby allowing the passage of the oligosaccharide moiety through the transporter while the polyprenyl tail remains partially embedded in the lipid bilayer but attached to the surface of PglK. This newly proposed mechanism is distinct from the alternating-access model that underlies the mechanism of most primary and secondary transporters studied so far.

PglK _in vitro_ LLO flipping and ATPase activity

Flipping of polyprenyl-linked oligosaccharides has previously been investigated using radiolabelled natural lipids[23,24]. By contrast, we exploited the fact that PglK can process LLOs with shortened oligosaccharides both _in vivo_ and _in vitro_[20,25,26]. We therefore generated proteoliposomes by co-reconstituting PglK and a trisaccharide LLO (tLLO), which contained a reducing-end bacillosamine and two N-acetylgalactosamine (GalNAc) moieties (Fig. 1a and Extended Data Fig. 1). To determine the rate of ATP-driven, PglK-dependent tLLO flipping, the amount of tLLO molecules in the outer, accessible leaflet of the proteoliposomes was quantified by radiolabelling using purified glycosyltransferase PglH[27–29], an enzyme of the _C. jejuni_ N-glycosylation machinery that was shown to attach up to three GalNAc residues to the non-reducing end of tLLO (Extended Data Fig. 2). We used excess UDP-GalNAc, thereby forcing three GalNAc molecules to be added to each tLLO molecule, allowing accurate determination of tLLO flipping to the liposome lumen (Fig. 1b and Extended Data Fig. 9a). The conversion of tLLO to hexa-saccharide LLO was confirmed by transferring the resulting oligosaccharides to fluorescently labelled acceptor peptides using purified oligosaccharyltransferase PglB, followed by SDS–PAGE analysis[26,30,31]. PglK-catalysed _in vitro_ tLLO flipping was strictly dependent on ATP hydrolysis. A PglK

[1]Institute of Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zürich, Switzerland. [2]Department of Chemistry and Biochemistry, University of Berne, CH-3012 Berne, Switzerland. [3]Institute of Microbiology, ETH Zürich, CH-8093 Zürich, Switzerland. †Present address: GlycoVaxyn AG, Grabenstrasse 3, 8952 Schlieren, Switzerland.
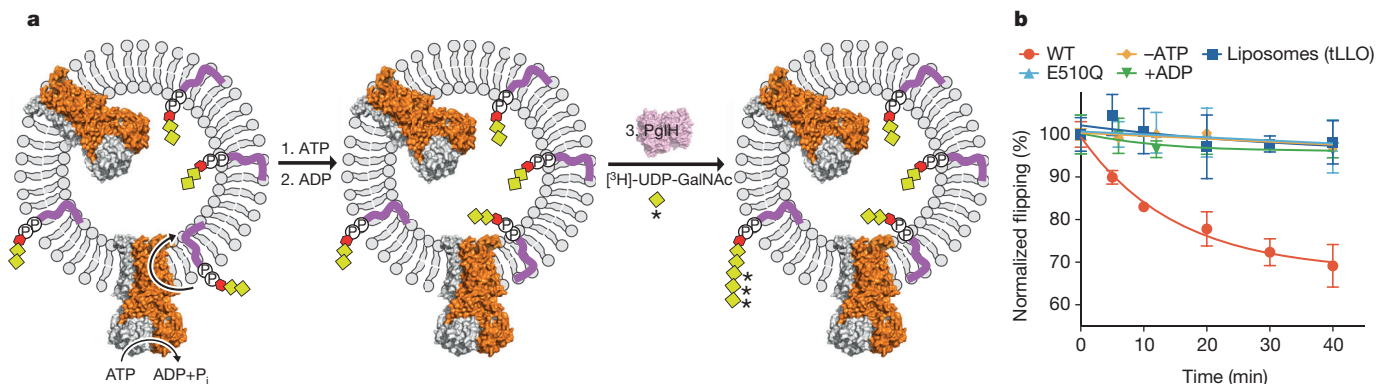
Figure 1 | *In vitro* LLO flipping assay. a, Schematic of designed assay. Proteoliposomes contain PglK homodimer (subunits in orange and grey) and tLLO, both in 50_{in}:50_{out} orientations (see Methods and Extended Data Fig. 9c, d). Owing to the external addition of ATP, only inside-out transporters catalyse tLLO flipping. The undecaprenyl tail of tLLO is shown in purple. Red hexagon denotes di-*N*-acetylbacillosamine; yellow square denotes *N*-acetylgalactosamine; and yellow square with an asterisk denotes ^{3}H-*N*-acetylgalactosamine. PglK activity is inhibited by the addition of excess ADP, after which proteoliposomes are incubated with purified PglH and ^{3}H-UDP-GalNAc to label and quantitate tLLO remaining in the external leaflet. **b**, tLLO flipping of wild-type (WT) PglK and the E510Q mutant. Controls include empty liposomes containing only tLLO (liposomes), the constant presence of 4 mM ADP (+ADP), and the absence of ATP (−ATP). Error bars denote s.d. ($n = 3$).

mutant containing a glutamine residue in the Walker-B motif (E510Q) exhibited an almost 100-fold decrease in the flipping rate, which correlated with strongly reduced *in vivo* activity (Extended Data Fig. 3a–c). The ATPase activity of PglK in the presence of LLO follows Michaelis–Menten kinetics, with a K_m value of 0.46 ± 0.08 mM (mean $\pm$ s.d.), matching the K_m of ATP-dependent tLLO flipping of 0.36 ± 0.03 mM (Extended Data Fig. 3d). This corroborates the interpretation that ATP hydrolysis is required and driving PglK-catalysed LLO flipping. As was observed with other ABC transporters, binding of native substrate can stimulate the ATPase activity. PglK-catalysed ATP hydrolysis is stimulated ~2.5-fold by full-length LLO and tLLO both in detergent and in liposomes, demonstrating that the terminal GalNAc and the branching glucose residues of native LLO are not relevant for ATPase stimulation or flipping (Extended Data Fig. 3e–g). Notably, although the hydrolysis and flipping rates are almost 100-fold lower than those of wild-type PglK, the ATPase activity of the E510Q mutant is also stimulated by LLO, suggesting that the interaction with LLO is not affected by the E510Q mutation. Unlike remotely related multi-drug ABC transporters[32,33], the stimulation of the ATPase activity of PglK is highly specific. Commonly used drugs such as verapamil, Hoechst33342, rhodamine 6G and acrilflavin had no stimulating effect (Extended Data Fig. 3f). The addition of lipid A, known to stimulate MsbA[34], was equally unable to stimulate the ATPase activity of PglK.

Structures of PglK in distinct states

We determined three crystal structures of the PglK mutant E510Q at resolutions ranging from 2.9 Å to 5.9 Å (Fig. 2a, Extended Data Fig. 4 and Extended Data Table 1). Two structures of the apo protein were determined at higher resolution and revealed distinct inward-facing conformations (designated apo-inward-1 and -2), whereas the third was of ADP-bound PglK, resulting in an unprecedented, outward-occluded conformation. Given the limited resolution of the latter, model building was helped by a previous finding[35] that transmembrane helices of ABC exporters move in pairs during structural transitions. We also confirmed the register of the resulting model by collecting anomalous diffraction data of Hg-soaked crystals and of a selenomethionine derivative, which allowed us to position cysteine and methionine residues (Extended Data Fig. 5).

The fold of PglK is similar to that of other ABC exporters, first revealed by the structure of the bacterial ABC transporter Sav1866 (ref. 36) (Extended Data Fig. 6a). However, PglK contains an additional, short α-helix (termed external helix 'EH') at the periplasmic membrane boundary and parallel to the plane of the membrane (see below). The apo-inward-1 and outward-occluded structures were obtained from PglK, crystallized using lauryl maltose neopentyl glycol (LMNG) as a detergent, whereas that of apo-inward 2 was obtained using *N*-dodecyl-β-D-maltopyranoside (DDM) (Extended Data Table 1).
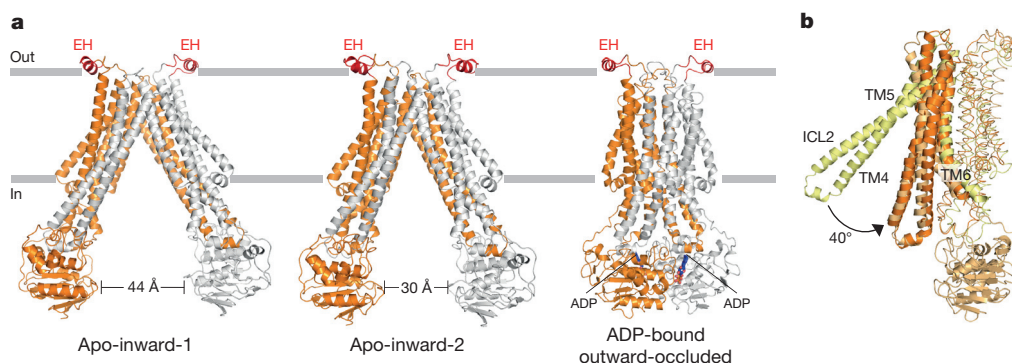

Figure 2 | Structures of PglK in distinct conformations. a, PglK subunits are in orange and grey. EH denotes the external helix. **b**, Conformational changes associated with the conversion of inward- and outward-facing states. A single PglK subunit is shown in light orange, with transmembrane helices 4–6 shown in ribbon representation. The same helices of a superimposed PglK subunit in the apo-inward-1 state (light yellow) and in the outward-occluded state (dark orange) are shown as ribbons. ICL, intracellular loop.

Crystal contacts in the three crystal forms are distinct, as are the crystallization conditions (Extended Data Fig. 7). This finding suggests that the type of detergent and the lattice contacts, rather than intrinsic properties or dynamics of the protein, define the degree of nucleotide-binding domain (NBD) separation in the inward-facing conformations. By contrast, the NBD arrangement in the outward-occluded state resembles a 'closed sandwich dimer' observed in several nucleotide-bound structures of ABC transporters[17,36] and in various isolated ABC domains[37,38]. A comparison of the three PglK structures revealed that the main conformational changes involve a scissor-like motion, with transmembrane helices TM4 and TM5 moving as a rigid body towards TM6 and tilting by 16° (apo-inward-1 versus apo-inward-2) or by 40° (apo-inward-1 versus outward-occluded) (Fig. 2b and Extended Data Fig. 6b). At the periplasmic surface, the outward-occluded conformation of PglK differs from the fully outward-facing Sav1866 structure[36] in that only a small opening from the central translocation pathway exists. Outward-occluded PglK thus reflects an intermediate conformation between the occluded state of the antibacterial peptide ABC exporter McjD[39] and the fully outward-open Sav1866 (ref. 36) (Extended Data Fig. 6C).

Role of external helix EH in PglK–LLO interaction

The loop connecting transmembrane helices TM1 and TM2 of PglK contains the external helix EH (Figs 2a and 3a), the orientation of which relative to the transmembrane domain is defined by several hydrophobic, hydrophilic and cation-π interactions (Fig. 3a). As a consequence, two hydrophobic grooves are formed near the external membrane boundary, one between each EH and TM1 and TM2 of

PglK (Fig. 3a). Given its unusual location and surface, we investigated the functional relevance of the EH by studying PglK mutants *in vivo* and *in vitro*. We generated a deletion mutant (termed ΔEH) in which the entire helix (residues S46–P67) was replaced by a flexible linker containing eight residues. In addition, we generated a triple mutant converting prominently located tyrosine residues into alanines (Y50A/Y56A/Y63A, termed EHm1) and a double mutant of positively charged residues involved in stabilizing interactions (R53A/K55A, termed EHm2). When reconstituted in liposomes, the three EH mutants showed unaltered basal ATPase activity compared to the wild type. However, no stimulation of ATPase activity by LLO was observed, suggesting that the mutations had abolished the role of EH in linking LLO binding to ATPase stimulation (Fig. 3b). In detergent, the basal ATPase levels of the two EH mutants EHm1 and EHm2 were higher than that of wild type, whereas that of ΔEH was slightly lower. Notably, however, ATPase stimulation by LLO was absent for all three EH mutants in detergent as it was in liposomes. In agreement with the lack of LLO-stimulated ATPase activity, the *in vitro* flipping rate of the ΔEH mutant was too low to be accurately determined, whereas those of EHm1 and EHm2 were found to be between 15- and 33-fold lower than wild type (Fig. 3c and Extended Data Fig. 3b). Furthermore, the ΔEH mutant had completely lost *in vivo* activity, whereas the mutants EHm1 and EHm2 revealed only slightly lower *in vivo* output compared to wild type (Fig. 3d). It is important to note that reductions of the *in vivo* output of the biosynthetic glycosylation pathway as a consequence of a PglK mutation may not directly correlate with the reduction of *in vitro* flipping rates because the rate-limiting step of the *in vivo* process is unknown. A similar phenomenon was previously observed for mutants of PglB that exhibited reductions of *in vitro* glycosylation rates by two orders of magnitude or more, but showing only marginally lower *in vivo* output[21,26,30,31].

We speculated that the hydrophobic grooves formed near the EH (Fig. 3a) might contribute to the binding of LLO to PglK. Such an interaction could anchor the lipid tail of LLO and favour the positioning of the pyrophosphate moiety for entering a translocation cavity and thus initiate the flipping reaction. Because no structure of PglK with bound LLO could be determined, we investigated PglK activity using synthetic compounds representing structurally truncated versions of native LLO (Fig. 3e and Extended Data Fig. 1). Notably, none of the truncated LLOs stimulated the ATPase activity of PglK reconstituted in proteoliposomes, suggesting that the shorter polyprenyl tails (maximum of 20 carbon atoms in geranyl-geranyl-PP-GlcNac) may prevent their proper recognition by PglK in a lipid bilayer (Extended Data Fig. 8). By contrast, the ATPase activity of PglK in detergent was stimulated by farnesyl-pyrophosphate (FPP), but also by farnesyl-PP-GlcNAc and geranyl-geranyl-PP-GlcNac. Unlike in lipid bilayers, the stimulation of the ATPase activity of PglK in detergent may be due to the simultaneous binding of several truncated LLO molecules. Ubiquinone, a molecule with a polyprenyl tail similar to that of native LLO but containing a distinct head-group (Fig. 3e and Extended Data Fig. 1), was unable to stimulate the ATPase activity of PglK both in detergent and in liposomes. This demonstrated that both the polyprenyl tail and the pyrophosphate moieties are essential for a productive, ATPase-stimulating interaction of LLO with PglK. When using native LLO, we found positive cooperativity of ATPase stimulation ($K_{m(LLO)} = 8.2 \pm 0.8\ \mu M$; Hill coefficient ($n_H$) = 1.9 ± 0.4, Extended Data Fig. 3h), which correlates with the presence of two regions on the surface of PglK that can interact with LLO molecules.

Functions of inward- and outward-facing cavities

The inward-facing PglK structures revealed a large cavity that is accessible to the cytoplasmic side and to the inner leaflet of the bilayer (Fig. 4a). Analogous cavities have been observed in nucleotide-free, inward-facing states of other ABC transporters[17,33,40], where they have been interpreted to reflect substrate-binding pockets. Accordingly, one might speculate that during LLO flipping, the 55-carbon polyprenyl tail
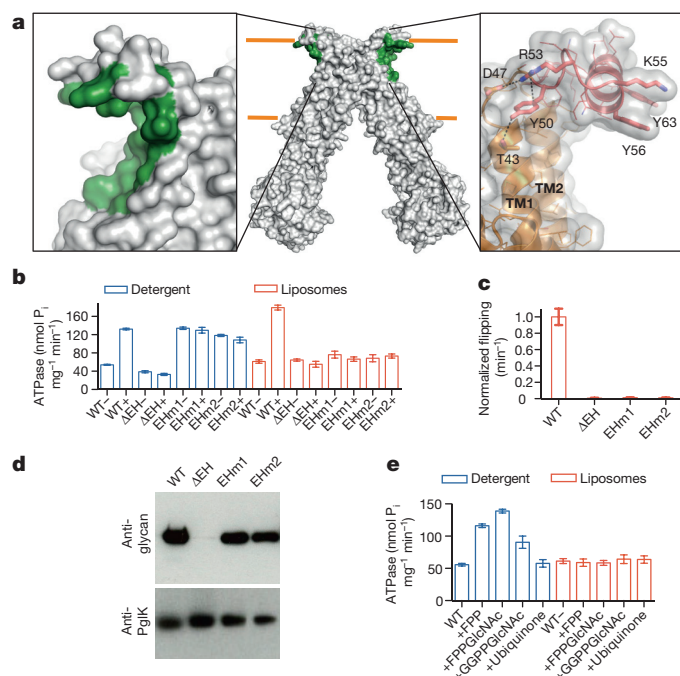

Figure 3 | Structural and biochemical characterization of the external helix EH. a, Left, hydrophobic groove (green) formed by the EH and transmembrane helices 1 and 2 (TM1 and TM2). Right, close-up of the EH with functionally relevant residues indicated. **b**, ATPase activity in the presence (+) or absence (−) of LLO of wild-type PglK and EH mutants in detergent (LMNG) or proteoliposomes. **c**, *In vitro* tLLO flipping rate of PglK variants. Bars indicate initial rates. **d**, *In vivo* LLO flipping by PglK variants expressed in *Escherichia coli* SCM6 cells. **e**, ATPase activity of PglK in detergent or proteoliposomes in the presence of diverse lipids and glycolipids: FPP (90 μM farnesyl-pyrophosphate), FPPGlcNAc (90 μM farnesyl-PP-GlcNAc), GGPPGlcNAc (90 μM geranyl-geranyl-PP-GlcNAc), and ubiquinone (90 μM). ΔEH, EH replaced by GSSGSSGS linker; EHm1, Y50A/Y56A/Y63A; EHm2, R53A/K55A. Error bars denote s.d. (*n* = 3).
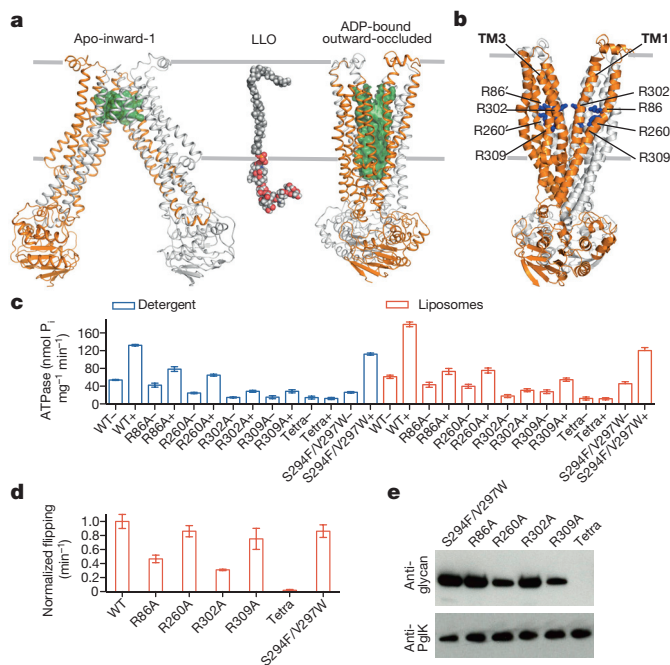
Figure 4 | Analysis of translocation pathway. a, Comparison of cavities (green) in inward-facing and outward-occluded states of PglK, with native LLO shown as space-filling model for size reference. **b**, Side view of fully outward-facing homology model of PglK based on Sav1866 structure[36]. Cavity-exposed, positively charged side chains are shown as blue sticks and labelled. **c**, ATPase activity of wild-type PglK and mutants in detergent (LMNG) and in proteoliposomes, in the presence (+) or absence (−) of native LLO. **d**, In vitro tLLO flipping rate of PglK variants. Bars indicate initial rates. **e**, In vivo LLO flipping of PglK variants expressed in E. coli SCM6 cells. Tetra, R86A/R260A/R302A/R309A. Error bars denote s.d. ($n = 3$).

enters the inward-facing cavity from the lipid bilayer to fold ('curl up') into the very bottom of the cavity, where the surface is hydrophobic (Fig. 4a and Extended Data Fig. 6d). Although difficult to reconcile with the essential nature of the EH in facilitating LLO interactions (see above), we tested this hypothesis and reduced the volume of the hydrophobic section of the inward-facing cavity by ∼60% (ref. 41) by introducing four bulky residues, two in each PglK subunit (S294F/V297W). The resulting mutant retained wild-type in vivo and in vitro flipping rates and unaltered ATPase stimulation (Fig. 4c–e), arguing against the involvement of this cavity in polyprenyl binding and suggesting that the polyprenyl tail remains exposed to the lipid bilayer during flipping.

We next considered whether the pyrophosphate-oligosaccharide head group could be transferred to the inward-facing cavity during flipping. However, this seems highly unlikely for the following reason: as a consequence of the domain-swapped architecture of ABC exporters, TM4–TM5 of each PglK subunit cross over and bind the NBD of the opposite subunit. The cavity formed by the inward-facing state is thus defined by the separation of TM4 and TM6 (Fig. 2b). If the pyrophosphate and oligosaccharide moieties of LLO were to enter this cavity, the LLO molecule would be 'pinched off' during the transition of PglK to outward-facing, as the gap between TM4 and TM6 closes and a distinct gap (between TM1 and TM3) opens to the outside. This would trap the LLO molecule and produce a massive steric clash, preventing substrate release. The above considerations lead to the inevitable conclusion that during the flipping cycle of PglK, the inward-facing cavity does not contribute to LLO binding.

The outward-facing cavity is distinct in chemical nature from the inward-facing one. Exploiting previous findings[35], we modelled a fully outward-facing state of PglK based on the structure of Sav1866 (ref. 36). The resulting model features a similar cavity as that of our outward-occluded structure, but with a wider opening to the periplasmic

side of the membrane (Fig. 4a, b). In both states, the outward-facing cavity spans almost the entire membrane, and whereas the outward-occluded conformation would not provide sufficient space to contain the pyrophosphate and oligosaccharide moieties of LLO, the fully outward-facing conformation could. We identified multiple arginine residues (R86, R260, R302 and R309) that lined the interior of the outward-facing cavity, accounting for a total of eight positive charges (Fig. 4b). Owing to the scissoring motion required to convert ABC exporters from inward- to outward-facing states, these arginines are buried and/or form salt-bridges in the inward-facing conformations but become accessible when PglK is in an outward-facing or outward-occluded state. Because arginines might serve as binding partners of the pyrophosphate moiety of LLO, we generated individual alanine mutants (R86A, R260A, R302A and R309A) and a fourfold mutant ('tetra') with all charges removed. Whereas the single mutants maintained a certain level of flipping activity both in vivo and in vitro, the fourfold mutant was unable to catalyse LLO flipping (Fig. 4d, e). Furthermore, whereas the single mutants retained some stimulation of ATPase activity by LLO (albeit with a lower basal ATPase rate), the ATPase activity of the fourfold mutant was no longer stimulated (Fig. 4c), suggesting an apparent loss of interaction between LLO and mutant PglK.

Flipping mechanism

The accumulated biochemical and structural evidence allow us to propose a mechanism of PglK-catalysed LLO flipping, shown in Fig. 5. The two conformations of PglK that are relevant for activity are the outward-occluded (ADP-bound) and outward-facing (ATP-bound) ones. Given millimolar concentrations of ADP and ATP in the cell, an apo-state of the transporter, represented by an inward-facing structure, is probably transient and sparsely populated. It is indeed conceivable that fast exchange of ATP for ADP would not allow the ATP-binding sites of homodimeric PglK protein to be nucleotide-free at the same time. We propose that while the polyprenyl tail of LLO interacts with the surface of PglK including and around the EH, the pyrophosphate moiety is directly transferred into the outward-facing cavity of the ATP-bound state, where it can form electrostatic interactions with exposed arginine residues. The oligosaccharide then follows the pyrophosphate moiety but does not require any specific interactions with PglK during the translocation process, which could explain the low specificity of the flipping reaction with respect to the type and number of saccharides. This hypothesis is consistent with the observation that PglK can flip distinct LLOs that are normally transported by the multidrug/oligosaccharidyl-lipid/polysaccharide (MOP)-type flippase Wzx[20]. It can also suggest how related ABC-type flippases could translocate long, linear polysaccharides during O-antigen generation[42]: as the elongated oligosaccharide chain is imagined to slide directly into an outward-facing translocation pathway and leave on the external side, no limitation on chain length may exist. In our model, the driving force of the flipping reaction of PglK has two components. First, as ATP hydrolysis causes the NBDs to be pushed apart for product release, transmission of this motion to the TMDs may cause pressure to be exerted on the substrate in the translocation pathway. This is reminiscent of the peristaltic component proposed for the mechanism of the vitamin B12 ABC importer[43]. Second, once the pyrophosphate and oligosaccharide moieties have diffused out of the translocation pathway, PglK will probably adopt an outward-occluded conformation as observed in our ADP-bound structure, thereby effectively closing the door behind the substrate. We propose that the energy gained by the hydrolysis of ATP is primarily used to break the interaction of the pyrophosphate moiety with the essential arginine residues lining the outward-facing cavity of PglK. It is conceivable that pyrophosphate-containing LLO accumulated on the periplasmic side might act as an inhibitor of the flipping reaction, which might contribute to the slowing of the in vitro flipping rate with time. In vivo, a
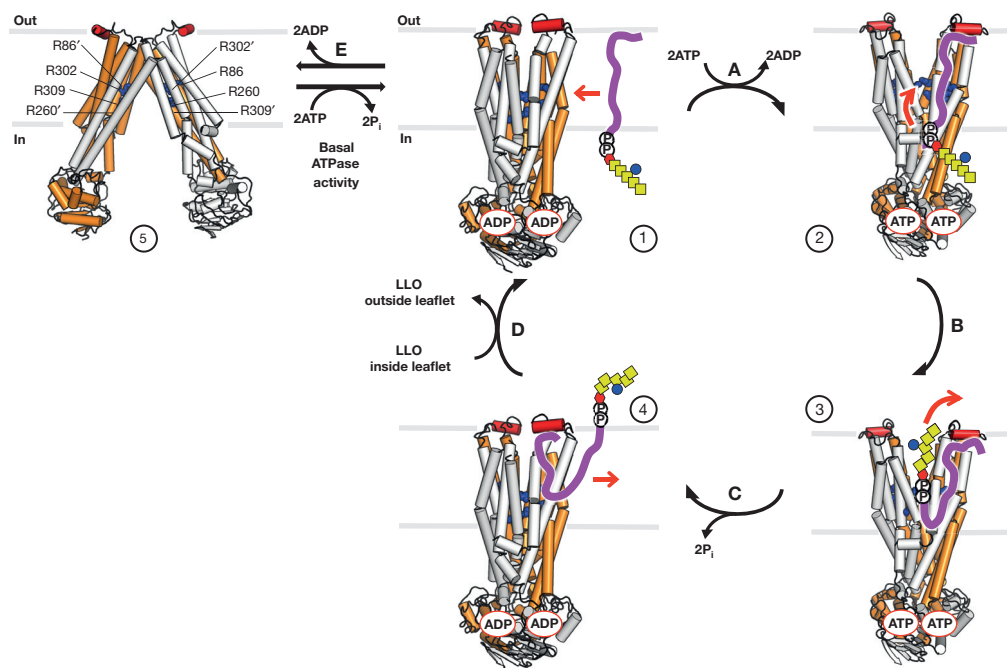
Figure 5 | Proposed LLO flipping mechanism including five states (circled numbers). PglK is shown as a cylinder model, with subunits coloured grey and orange, and the EH coloured red. Functionally important arginine residues are shown as blue sticks and labelled. The polyprenyl tail of LLO is shown in purple; phosphate groups are indicated by circled 'P'. Red hexagon denotes di-N-acetylbacillosamine; yellow square denotes N-acetylgalactosamine; and blue circle denotes glucose. State 5 reflects the apo-inward-1 structure, states 1 and 4 are based on the outward-occluded structure, states 2 and 3 are modelled outward-open conformations. The proposed molecular events indicated by arrows are: A, polyprenyl tail of LLO interacting with PglK region including and around EH, ATP replacing ADP; B, pyrophosphate-oligosaccharide head group entering outward-facing cavity; C, ATP hydrolysis, LLO head-group release on external side of the membrane, release of inorganic phosphate; D, polyprenyl tail release; E, futile ATPase cycle.

similar problem might occur if undecaprenyl-pyrophosphate were to accumulate after glycan transfer by PglB[21], but is probably alleviated by the catalysed hydrolysis of undecaprenyl-pyrophosphate to the monophosphate species[22] (Extended Data Fig. 2).

The LLO flipping mechanism deduced for PglK has a resemblance to the 'credit card swipe' model proposed for the flipping of phospholipids by P4-ATPases flippases[44–46]. Two key differences are the proposed recognition of the polyprenyl tail on the PglK surface by means of the EH, and the requirement of a long, sufficiently wide translocation pathway that contains the pyrophosphate and oligosaccharide moieties during flipping. Our mechanism is distinct from that proposed for MsbA function, which suggested that the entire lipid A-core may enter the inward-facing, nucleotide-free state of the transporter[17]. More generally, our mechanistic proposal is also distinct from the canonical alternating-access model used to explain the mechanism of transporters of small, hydrophilic substrates.

Conclusions

Our results define the structural and mechanistic basis of lipid-linked oligosaccharide flipping by an ABC transporter. The data suggest a flipping mechanism suitable for the translocation of a large head group containing a pyrophosphate moiety and a linear polysaccharide, both of which are shielded from the lipid bilayer in a translocation cavity. We predict that the proposed direct jump into the outward-facing conformation and the putative interaction of the lipid tail with the membrane-facing surface of the transporter may be a feature of other ABC-type LLO flippases and possibly of MOP-type transporters involved in bacterial cell wall assembly or N-glycan flipping in eukaryotes[14,20]. Understanding the molecular mechanism of LLO flipping may open new ways to modify the glycosylation machinery in bacteria, which could be exploited for the generation of new glycoconjugates or glycoproteins that may have applications in diagnosis or therapeutics[47,48].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta* **1426,** 239–257 (1999).
2. Helenius, J. *et al.* Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* **415,** 447–450 (2002).
3. Sprong, H., van der Sluijs, P. & van Meer, G. How proteins move lipids and lipids move proteins. *Nature Rev. Mol. Cell Biol.* **2,** 504–513 (2001).
4. Sebastian, T. T., Baldridge, R. D., Xu, P. & Graham, T. R. Phospholipid flippases: building asymmetric membranes and transport vesicles. *Biochim. Biophys. Acta* **1821,** 1068–1077 (2012).
5. Hankins, H. M., Baldridge, R. D., Xu, P. & Graham, T. R. Role of flippases, scramblases and transfer proteins in phosphatidylserine subcellular distribution. *Traffic* **16,** 35–47 (2015).
6. Krahling, S., Callahan, M. K., Williamson, P. & Schlegel, R. A. Exposure of phosphatidylserine is a general feature in the phagocytosis of apoptotic lymphocytes by macrophages. *Cell Death Differ.* **6,** 183–189 (1999).
7. Balasubramanian, K. & Schroit, A. J. Aminophospholipid asymmetry: A matter of life and death. *Annu. Rev. Physiol.* **65,** 701–734 (2003).
8. Cuthbertson, L., Kos, V. & Whitfield, C. ABC transporters involved in export of cell surface glycoconjugates. *Microbiol. Mol. Biol. Rev.* **74,** 341–362 (2010).
9. Cuthbertson, L., Kimber, M. S. & Whitfield, C. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proc. Natl Acad. Sci. USA* **104,** 19529–19534 (2007).
10. Clarke, B. R., Cuthbertson, L. & Whitfield, C. Nonreducing terminal modifications determine the chain length of polymannose O antigens of *Escherichia coli* and couple chain termination to polymer export via an ATP-binding cassette transporter. *J. Biol. Chem.* **279,** 35709–35718 (2004).
11. Sharom, F. J. Flipping and flopping–lipids on the move. *IUBMB Life* **63,** 736–746 (2011).
12. Kodigepalli, K. M., Bowers, K., Sharp, A. & Nanjundan, M. Roles and regulation of phospholipid scramblases. *FEBS Lett.* **589,** 3–14 (2015).
13. Brunner, J. D., Lim, N. K., Schenck, S., Duerst, A. & Dutzler, R. X-ray structure of a calcium-activated TMEM16 lipid scramblase. *Nature* **516,** 207–212 (2014).
14. Hvorup, R. N. *et al.* The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *Eur. J. Biochem.* **270,** 799–813 (2003).
15. Lopez-Marques, R. L., Theorin, L., Palmgren, M. G. & Pomorski, T. G. P4-ATPases: lipid flippases in cell membranes. *Pflugers Arch.* **466,** 1227–1240 (2014).

16. Eckford, P. D. & Sharom, F. J. The reconstituted *Escherichia coli* MsbA protein displays lipid flippase activity. *Biochem. J.* **429,** 195–203 (2010).

17. Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl Acad. Sci. USA* **104,** 19005–19010 (2007).

18. Wacker, M. *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli. Science* **298,** 1790–1793 (2002).

19. Young, N. M. *et al.* Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni. J. Biol. Chem.* **277,** 42530–42539 (2002).

20. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25,** 967–976 (2006).

21. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).

22. Tatar, L. D., Marolda, C. L., Polischuk, A. N., van Leeuwen, D. & Valvano, M. A. An *Escherichia coli* undecaprenyl-pyrophosphate phosphatase implicated in undecaprenyl phosphate recycling. *Microbiology* **153,** 2518–2529 (2007).

23. Sanyal, S. & Menon, A. K. Specific transbilayer translocation of dolichol-linked oligosaccharides by an endoplasmic reticulum flippase. *Proc. Natl Acad. Sci. USA* **106,** 767–772 (2009).

24. Sanyal, S. & Menon, A. K. Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase. *Proc. Natl Acad. Sci. USA* **107,** 11289–11294 (2010).

25. Linton, D. *et al.* Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol.* **55,** 1695–1703 (2005).

26. Lizak, C. *et al.* A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *J. Biol. Chem.* **289,** 735–746 (2014).

27. Troutman, J. M. & Imperiali, B. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry* **48,** 2807–2816 (2009).

28. Abeijon, C. & Hirschberg, C. B. Topography of initiation of *N*-glycosylation reactions. *J. Biol. Chem.* **265,** 14691–14695 (1990).

29. Hanover, J. A. & Lennarz, W. J. The topological orientation of *N,N′*-diacetylchitobiosylpyrophosphoryldolichol in artificial and natural membranes. *J. Biol. Chem.* **254,** 9237–9246 (1979).

30. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: *in vitro* quantification of sequon binding and catalysis. *J. Biol. Chem.* **288,** 8849–8861 (2013).

31. Lizak, C. *et al.* Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nature Commun.* **4,** 2627 (2013).

32. Siarheyeva, A. & Sharom, F. J. The ABC transporter MsbA interacts with lipid A and amphipathic drugs at different sites. *Biochem. J.* **419,** 317–328 (2009).

33. Jin, M. S., Oldham, M. L., Zhang, Q. & Chen, J. Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans. Nature* **490,** 566–569 (2012).

34. Raetz, C. R., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* **76,** 295–329 (2007).

35. Lee, J. Y., Yang, J. G., Zhitnitsky, D., Lewinson, O. & Rees, D. C. Structural basis for heavy metal detoxification by an Atm1-type ABC exporter. *Science* **343,** 1133–1136 (2014).

36. Dawson, R. J. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443,** 180–185 (2006).

37. Zaitseva, J. *et al.* A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. *EMBO J.* **25,** 3432–3443 (2006).

38. Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell* **10,** 139–149 (2002).

39. Choudhury, H. G. *et al.* Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. *Proc. Natl Acad. Sci. USA* **111,** 9145–9150 (2014).

40. Shintre, C. A. *et al.* Structures of ABCB10, a human ATP-binding cassette transporter in apo- and nucleotide-bound states. *Proc. Natl Acad. Sci. USA* **110,** 9710–9715 (2013).

41. Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38,** W555–W562 (2010).

42. Whitfield, C. & Trent, M. S. Biosynthesis and export of bacterial lipopolysaccharides. *Annu. Rev. Biochem.* **83,** 99–128 (2014).

43. Korkhov, V. M., Mireku, S. A. & Locher, K. P. Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F. *Nature* **490,** 367–372 (2012).

44. Vestergaard, A. L. *et al.* Critical roles of isoleucine-364 and adjacent residues in a hydrophobic gate control of phospholipid transport by the mammalian P4-ATPase ATP8A2. *Proc. Natl Acad. Sci. USA* **111,** E1334–E1343 (2014).

45. Stone, A. & Williamson, P. Outside of the box: recent news about phospholipid translocation by P4 ATPases. *J. Chem. Biol.* **5,** 131–136 (2012).

46. Pomorski, T. & Menon, A. K. Lipid flippases and their biological functions. *Cell. Mol. Life Sci.* **63,** 2908–2921 (2006).

47. Hug, I. & Feldman, M. F. Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology* **21,** 138–151 (2011).

48. Jones, C. Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *An. Acad. Bras. Cienc.* **77,** 293–324 (2005).

Author Contributions C.P. determined the structures of PglK, established the *in vitro* flipping assay, and performed *in vivo* flipping studies. S.G. crystallized PglK in the apo-inward-2 state, M.B. assisted in expression and purification of PglK. J.B., T.D. and J.-L.R. synthesized LLO analogues. K.P.L., S.G. and C.P. conceived the project. K.P.L., M.A., and C.P. analysed the data. K.P.L. and C.P. wrote the manuscript.

Author Information Atomic coordinates and structure factors were deposited with the RCSB Protein Data Bank (PDB) under accessions 5C78 (apo-inward-1), 5C76 (apo-inward-2) and 5C73 (outward-occluded). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.P.L. (locher@mol.biol.ethz.ch).

METHODS

No statistical methods were used to predetermine sample size. Experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

PglK expression and purification. The gene encoding *C. jejuni* PglK was cloned into a modified pET-19b vector (Novagen) with a His$_{10}$ affinity tag fused to N terminus. PglK was overexpressed in *E. coli* BL21-Gold (DE3) (Stratagene) cells in a 30 l fermentor (Infors). Cells were grown at 37 °C in modified Terrific Broth (TB) medium supplemented with 1% glucose (w/v) to $A_{600\,nm}$ of 10.0 before expression was induced by the addition of 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1 h. All following steps were performed at 4 °C unless specified differently. Cells were collected by centrifugation, re-suspended in 50 mM Tris-HCl, pH 8.0; 500 mM NaCl; 7 mM β-mercaptoethanol; 0.5 mM phenyl-methanesulfonylfluoride (PMSF). They were disrupted in a M-110L microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure. Membranes were pelleted by ultracentrifugation at 100,000*g* for 0.5 h. PglK was solubilized in 50 mM Tris-HCl, pH 8.0; 500 mM NaCl; 20 mM imidazole; 15% glycerol (v/v); 7 mM β-mercaptoethanol; 1% DDM (w/v) (Anatrace); 1% C$_{12}$E$_8$ anapoe (Anatrace) by stirring for 2 h. The supernatant was loaded onto a NiNTA superflow affinity column (Qiagen), washed once with the same buffer but containing 10% glycerol (v/v), 50 mM imidazole and 0.016–0.03% DDM, and then washed a second time with the same buffer containing 0.016–0.03% LMNG (Affymetrix). Elution was performed in the same buffer but containing 200 mM NaCl and 200 mM imidazole. The protein was further purified by size exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare) and peak fractions were pooled and concentrated to 8–10 mg ml^{-1} in an Amicon Ultra-15 concentrator (Millipore) with a molecular mass cut-off of 100 kDa.

Seleno-methionine derivative production. PglK was overexpressed in *E. coli* BL21 (DE3) RIPL (Stratagene) cells as described above with minor modifications. In brief, cells in TB-glucose media were grown at 37 °C to $A_{600\,nm}$ of 0.5–1.0. The cells were then used to inoculate a culture of M9 media supplemented with vitamin B1 hydrochloride. Cells were then grown until $A_{600\,nm}$ of 0.5 and used to inoculate 2 l M9 media supplemented with vitamin B1 hydrochloride. Cells were grown overnight at 37 °C until $A_{600\,nm}$ of ~0.9. At that point a cocktail of amino acids (lysine, threonine, phenylalanine, leucine, isoleucine and valine, all at 100 mg l^{-1}) including selenomethionine (200 mg l^{-1}) was added to the culture. After 30 min, expression was induced with 0.2 mM IPTG for 90 min. Cells were collected immediately thereafter.

Native crystals. For crystallization of the outward-occluded state PglK was incubated for 0.5 h with 5–10 mM ADP and 5–10 mM MgCl$_2$. PglK was crystallized by vapour diffusion in sitting drops or hanging drops at 20 °C using reservoirs containing 100 mM Tris-HCl, pH 8.2; 150 mM Mg-acetate; 25% PEG400 for the apo-inward-1 state crystallization; 50 mM glycine-NaOH, pH 9.4, 300 mM KCl, 50 mM MgCl$_2$, 21% PEG600 for the apo-inward-2 state crystallization; and 50 mM Na/K phosphate, pH 7.5, 250 mM NaCl, 3.5 M ammonium sulphate for the outward-occluded state crystallization. The protein to reservoir volume ratio was 2:1–1:1. Crystals typically appeared after 3–4 days and matured to full size within 2 weeks. Crystals were cryoprotected by gently increasing the cryoprotectant concentration in the drops (up to 30% PEG400 for apo-inward-1 and -2, and up to 3.0 M ammonium sulphate for outward-occluded) and directly flash frozen by immersion in liquid nitrogen before data collection.

Heavy-metal derivatives. Native crystals were soaked for 1 day in 1 mM ethyl mercury phosphate (EMP) or 1 mM potassium tetrachloroplatinate before back-soaking and flash-freezing by immersion in liquid nitrogen.

Data collection. Crystals belonged to the space groups *P*1 (apo-inward-1), *P*2$_1$ (apo-inward-2) and *P*4$_3$2$_1$2 (outward-occluded). Native data were collected at beamline X06SA at the Swiss Light Source (Villigen). EMP and PtCl$_4$ derivative data sets (Extended Data Table 1) were collected at the same station. Data were processed and merged with XDS[49] and anisotropic scaling/ellipsoid truncation was performed[50].

To improve the usable resolution and quality of the resulting electron density maps, we used the Karplus CC* (Pearson correlation coefficient)-based data cut-off approach[51] (Extended Data Table 1). The resolution limit was set taking into account a CC$_{1/2}$ > ~40% based on data merging statistics and a CC* analysis against unmerged intensities in the Phenix package[52], satisfying Karplus CC* against CC$_{work}$ and CC$_{free}$ criteria, as well as R_{free} of the highest resolution shell against the refined structure being less than or equal to ~50% (Extended Data Table 1).

Structure determination. Experimental phases of the apo-inward-1 structure were determined using multiple isomorphous replacement with anomalous scattering (MIRAS). Mercury positions were found using SHELX[53] and refined using SHARP[54]. Solvent flattening was performed using Solomon[55] and the resulting phases were combined with a higher-resolution native data set, followed by

non-crystallographic symmetry averaging and solvent flatting. Electron density maps were calculated using CCP4 programs[56]. The protein model was built using Coot[57], and refined using Phenix[52]. X-ray data and refinement statistics are given in Extended Data Table 1.

The apo-inward-2 structure was solved by molecular replacement using a partial model from the apo-inward-1 structure as a search model using the program Phaser[58], and refinement was performed using Phenix[52] combined with manual building in Coot[57]. The structure of the outward-occluded state was solved by a combination of molecular replacement using a partial model from the apo-inward-1 structure, and a single wavelength anomalous dispersion (SAD) data set from a selenomethionine derivative of PglK, as well as data from a derivative after labelling cysteine and histidine residues with EMP and PtCl4, respectively, in order to confirm the polypeptide chain register (Extended Data Fig. 5a–c). X-ray data and refinement statistics are given in Extended Data Table 1.

PglH expression and purification. The gene encoding *C. jejuni* PglH was cloned into a modified pET-19b vector (Novagen) with a N-terminal His$_{10}$ affinity tag fused to PglH. PglH was overexpressed in *E. coli* BL21-Gold (DE3) (Stratagene) cells in modified TB medium supplemented containing 1% glucose (w/v). Cells were grown at 37 °C to $A_{600\,nm}$ of 3.0 before the culture was induced by the addition of 0.5 mM IPTG, followed by a transfer to 18 °C for 16 h. All following steps were performed at 4 °C unless specified differently. Cells were collected by centrifugation, re-suspended in 50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM imidazole and 0.5 mM PMSF. Cells were disrupted in a M-110L microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure followed by the addition of 1% Triton X-100 (w/v)[27]. After centrifugation, the supernatant was loaded onto a NiNTA superflow affinity column (Qiagen), washed once with the same buffer but containing 50 mM imidazole. Elution was performed using a buffer containing 50 mM Tris-HCl, pH 8.0 and 500 mM imidazole. The protein was desalted into 50 mM Tris-HCl, pH 8.0 and 200 mM NaCl.

LLO and tLLO extraction. Isolation of LLOs was performed as described previously[30]. In brief, LLOs were extracted from *E. coli* SCM6 cells carrying a *C. jejuni* pglB$_{mut}$ cluster, containing an inactivated *pglB* gene (for LLO extraction) or a pglH$_{mut}$:pglB$_{mut}$ cluster, containing an additionally inactivated *pglH* gene (for tLLO extraction). Extraction was performed using a mixture of chloroform:-methanol:water (10:10:3) for LLO extraction or methanol:chloroform (1:2) for tLLO extraction. Extracts were dried in a rotary evaporator and reconstituted in a buffer containing 10 mM Tris, pH 8.0, 150 mM NaCl and 1% Triton X-100 (w/v). The concentration of reconstituted LLOs was determined by titrating various amounts of LLOs against a constant amount of acceptor peptide in an *in vitro* glycosylation assay as described before[59].

Reconstitution of PglK and variants in preoteoliposomes. Liposomes (20 mg lipid ml^{-1}) from a mixture 3:1 (w:w) of *E. coli* polar lipids and L-α-phosphatidilcho-line (Avanti Polar Lipids) were prepared by extrusion through polycarbonate filters (400 nm pore size) and diluted in 10 mM Tris-HCl, pH 8.0, 150 mM NaCl and 2 mM β-mercaptoethanol. After saturation with Triton X-100, the liposomes were mixed with purified protein at a lipid/protein ratio of 30–50:1 (w/w) and ~4 μM tLLO. BioBeads were then added to remove detergent. Finally, proteoliposomes containing a final concentration of 20 mg ml^{-1} lipids, 6.2–10.1 μM PglK and 20 μM tLLO were centrifuged and washed before being frozen in liquid nitrogen and stored at −80 °C.

Determination of PglK orientation in liposomes. A fully functional mutant of PglK containing none of the native cysteines but with an engineered cysteine at each NBD (C269L/C352S/C386S/C549L/S544C) was reconstituted in proteolipo-somes. After incubation with the thiol-reactive fluorescent dye Alexa Fluor 488 C5 maleimide, samples were analysed by SDS–PAGE and a fluorescence gel scanner. The ratio of PglK with NBDs facing outwards was calculated after comparison of the band intensities obtained from non-disrupted and disrupted proteoliposomes (0.3% Triton X-100). From this analysis it was determined that 51.8 ± 2.8% of PglK molecules are oriented with NBDs facing outwards.

***In vitro* tLLO flipping assay.** PglK proteoliposomes diluted in 10 mM Tris-HCl, pH 8.0, 150 mM NaCl and 2 mM β-mercaptoethanol were extruded through polycarbonate filters (400-nm pore size) and incubated with 5 mM MgCl$_2$ and 2 mM ATP to initiate the tLLO flipping reaction. To stop the translocation reaction, samples were diluted into a buffer containing 4 mM ADP. Labelling of non-flipped tLLO remnant in the external leaflet of proteoliposomes was achieved by incubation with purified glycosyltransferase PglH in the presence of 50 μM ^{3}H-UDP-GalNAc and 5 mM MnCl$_2$. To stop the labelling reaction, the samples were diluted into a cold stop buffer and filtered using a Multiscreen vacuum manifold (MSFBN6B filter plate, Millipore). Radioactivity trapped on the filters was determined using a gamma counter (CobraII Auto-Gamma, Packard). Nonlinear fitting of data and initial velocities determination were performed using GraphPad Prism 5 and the following equation: $Y = (Y0 - \text{plateau}) \times e^{(-K \times X)} + \text{plateau}$ Initial rates were calculated as the derivative at time = 0.

Product analysis and determination of tLLO orientation in liposomes. Glycosylation of tLLO by purified PglH was achieved after incubation of non-disrupted or disrupted proteoliposomes (0.3% Triton X-100) by addition of 3.5–50 μM UDP-GalNAc, 2 mM $MnCl_2$ and 0.2 mg ml^{-1} PglH. The reaction was stopped by heat shock (90 °C for 10 min). The products were then analysed by *in vitro* glycosylation of a fluorescently labelled substrate peptide (DQNAT sequon) catalysed by PglB[30,31]. In brief, the reaction products were incubated in a mixture containing 10 mM MES, pH 6.5, 100 mM NaCl, 10 mM $MnCl_2$, 3% glycerol (v/v), 1% Triton X-100 (w/v), 0.5 μM fluorescently labelled acceptor peptide and 1 μM purified PglB. The mixture was incubated at 30 °C and stopped by addition of SDS-containing gel loading buffer. Samples were analysed by Tricine–SDS–PAGE in mini-gels (8 × 8 cm) consisting of a 16% resolving gel with 6 M urea, a 10% spacer gel, and a 4% stacking gel[30,31]. Fluorescent bands for glycopeptide were visualized using a Typhoon Trio Plus imager (GE Healthcare) with excitation at 488 nm and a 526-nm SP emission filter. The amount of formed glycopeptide was determined from band intensities of fluorescence gel scans (ImageJ).

ATPase assays. ATP hydrolysis reactions were performed as described previously, using a modified molybdate-based colorimetric method[43]. Protein concentrations in the assays ranged from 0.1 mg ml^{-1} to 0.3 mg ml^{-1}. All reactions were performed in the presence of 2 mM ATP (or a range of ATP concentrations for K_m determination) and 5 mM $MgCl_2$. ATPase rates were determined using linear regression. Nonlinear regression and statistical analysis was performed using GraphPad Prism 5.

***In vivo* LLO flipping assay.** To analyse the flipping activity of PglK, we used an assay similar to that previously described to determine *in vivo* PglB glycosylation activity[30]. PglK and variants were cloned into a pMLBAD plasmid[60] with an N-terminal His$_{10}$ tag fused to it. This plasmid was transformed into *E. coli* cells defective for O-antigen biosynthesis (SCM6), carrying the plasmids pCL21 (ref. 59) and pACYC-wlaB::kan[25]. pCL21 encodes for the expression of the single-chain Fv fragment of 3D5 carrying a DQNAT glycosylation site in the linker region and a C-terminal Myc tag fused to 3D5. pACYC-wlaB::kan codes for the biosynthesis of the *C. jejuni* LLO with a knockout of the *C. jejuni* pglK gene. A 5 ml pre-culture was inoculated from a single clone and grown overnight at 37 °C in LB medium. The main culture was inoculated to an optical density $A_{600\,nm}$ of 0.05 in 10 ml LB medium and grown at 37 °C to $A_{600\,nm}$ of 0.5. The culture was induced by addition of arabinose to 0.1% (w/v) and grown for 4 h at 24 °C. For extraction of periplasmic proteins, an equivalent of 1 ml culture volume with an $A_{600\,nm}$ of 2 was collected by centrifugation, re-suspended in 100 μl extraction buffer, consisting of 30 mM Tris-HCl, pH 8.5, 20% (w/v) sucrose, 1 mM EDTA and 1 mg ml^{-1} lysozyme (Sigma) and incubated for 1 h at 4 °C. A final centrifugation step yielded periplasmic proteins in the supernatant. Glycosylation of 3D5 and expression of PglK were analysed by immunoblot following SDS–PAGE. Immunodetection was performed with anti-glycan serum hR6 (ref. 20) to observe glycosylated 3D5. Immunodetection of PglK was performed with anti-His-HRP serum (Sigma).

Mutagenesis. PglK mutants were generated by the QuickChange method or by using gBlocks gene fragments (Integrated DNA technologies). The resulting plasmids of all constructs were validated by DNA sequencing (Microsynth). PglK variants were cloned into pMLBAD as above and used for *in vivo* flipping assays.

Truncated LLOs synthesis. Farnesyl- and geranyl-geranyl-PP-GlcNAc were synthesized in seven steps from commercially available lipids and sugar according to the previously published strategy with slight modifications[61]. As the procedure involves the formation of highly acid sensitive phosphorylated intermediates, flash column chromatography were performed on silica gel basified with an ammonium hydroxide solution to avoid product degradation during purification. The final LLOs were obtained from the protected precursor by deacetylation using an ammonium hydroxide solution in methanol for 6 h. After freeze-drying, 16 mg of farnesyl-PP-GlcNAc and 42 mg of geranyl-geranyl-PP-GlcNAc (5% overall yield from the lipid) were respectively isolated as final compounds and were shown to be pure by ^{1}H, ^{13}C, ^{31}P NMR as well as negative-mode ESI-HRMS.

49. Kabsch, W. Xds. *Acta Crystallogr. D* **66,** 125–132 (2010).
50. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
51. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
52. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
53. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
54. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59,** 2023–2030 (2003).
55. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
56. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
57. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
58. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
59. Lizak, C., Fan, Y. Y., Weber, T. C. & Aebi, M. N-Linked glycosylation of antibody fragments in Escherichia coli. *Bioconjug. Chem.* **22,** 488–496 (2011).
60. Lefebre, M. D. & Valvano, M. A. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates. *Appl. Environ. Microbiol.* **68,** 5956–5964 (2002).
61. Liu, F. *et al.* Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation. *J. Am. Chem. Soc.* **136,** 566–569 (2014).
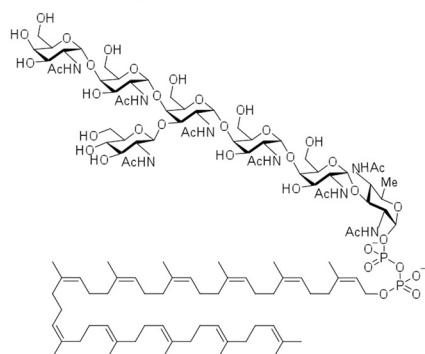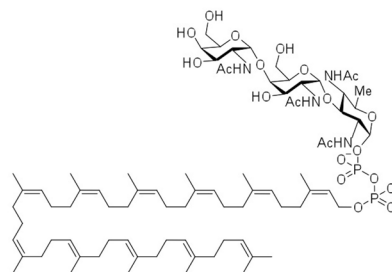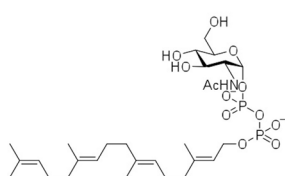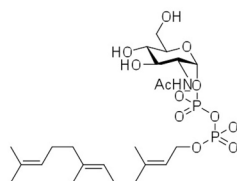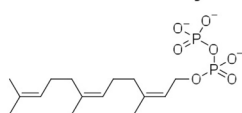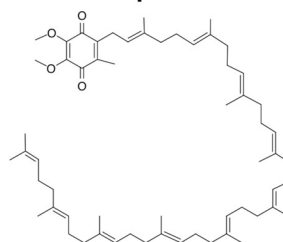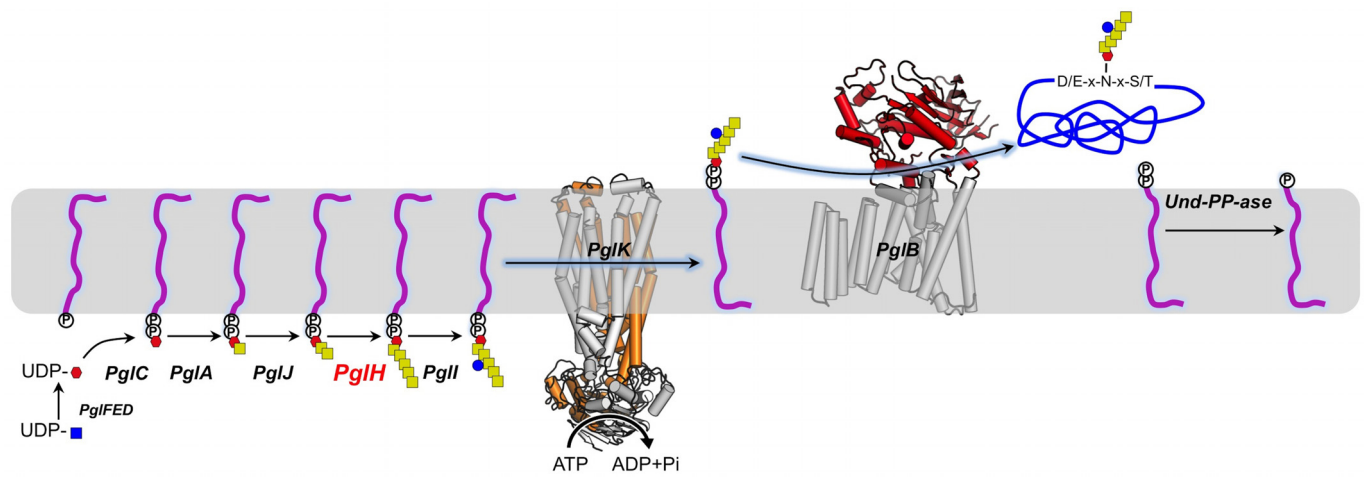
Extended Data Figure 1 | Structures of LLOs and chemical analogues used in this study.

Extended Data Figure 2 | LLO synthesis and *N*-glycosylation in *C. jejuni.* Red hexagon denotes di-*N*-acetylbacillosamine; yellow square denotes *N*-acetylgalactosamine; and blue circle denotes glucose. The LLO polyprenyl tail is shown in purple. A periplasmic polypeptide chain is shown in blue. PglH (red) is the glycosyltransferace used in the *in vitro* tLLO flipping assays. Und-PP-ase, undecaprenyl pyrophosphatase.
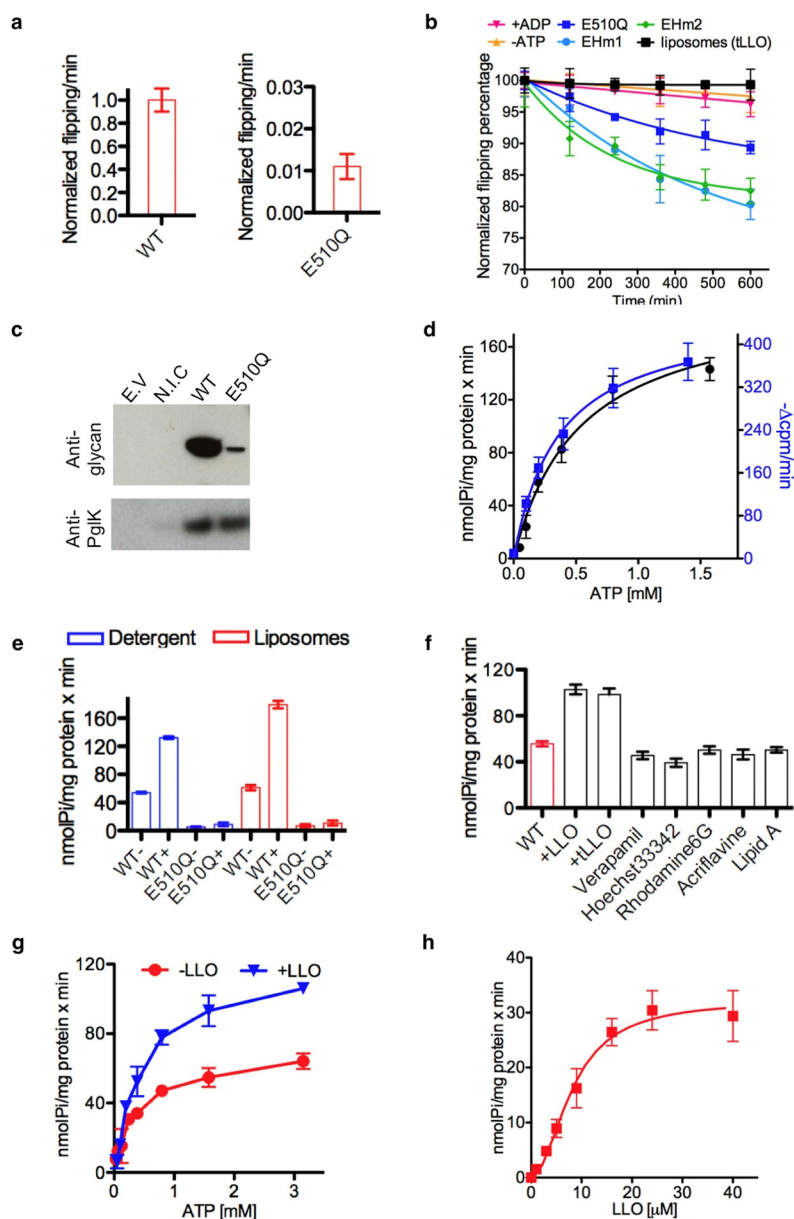
Extended Data Figure 3 | Biochemical characterization of wild-type and mutant PglK. a, *In vitro* tLLO flipping rates of PglK wild-type and E510Q mutant. Bars indicate initial velocities. **b**, *In vitro* tLLO flipping of PglK mutants E510Q, EHm1 and EHm2. +ADP, assay in the presence of 4 mM ADP; −ATP, assay in the absence of ATP. Liposomes, empty liposomes containing only tLLO. ADP alone does not cause tLLO to disappear from the external liposomes leaflet. **c**, *In vivo* LLO flipping of wild-type PglK and E510Q mutant expressed in *E. coli* SCM6 cells containing the *C. jejuni pgl* operon[30]. E.V., empty vector, N.I.C., non-induced cells. Anti-glycan refers to HR6 antibody recognizing the *N*-glycan of a substrate protein (see Methods), whereas anti-PglK is used to monitor PglK expression level. **d**, Determination of K_m values of ATP hydrolysis and ATP-driven *in vitro* tLLO flipping. The black curve represents the ATPase rate of PglK at distinct ATP concentrations. The blue curve

represents the initial LLO flipping rate in proteoliposomes at distinct ATP concentrations. Cpm, counts per million. **e**, ATPase activity in the presence (+) or absence (−) of native LLO of wild-type PglK and E510Q mutant in detergent (LMNG) or proteoliposomes. **f**, ATPase activity of PglK in detergent in the presence of LLO (20 μM), tLLO (20 μM), diverse drugs (50 μM) and lipid A (20 μM). **g**, Determination of K_m values of ATP hydrolysis in the presence (+) or absence (−) of LLO (20 μM). Stimulation results in higher V_{max}, while the K_m for ATP remains almost unaltered in the presence ($K_{m(+LLO)} = 0.54 \pm 0.03$ mM) or absence ($K_{m(-LLO)} = 0.36 \pm 0.04$ mM) of native LLO.
h, Determination of K_m value of ATPase stimulation by native LLO in detergent (rates are normalized against the basal activity in absence of LLO). Error bars denote s.d. ($n = 3$).
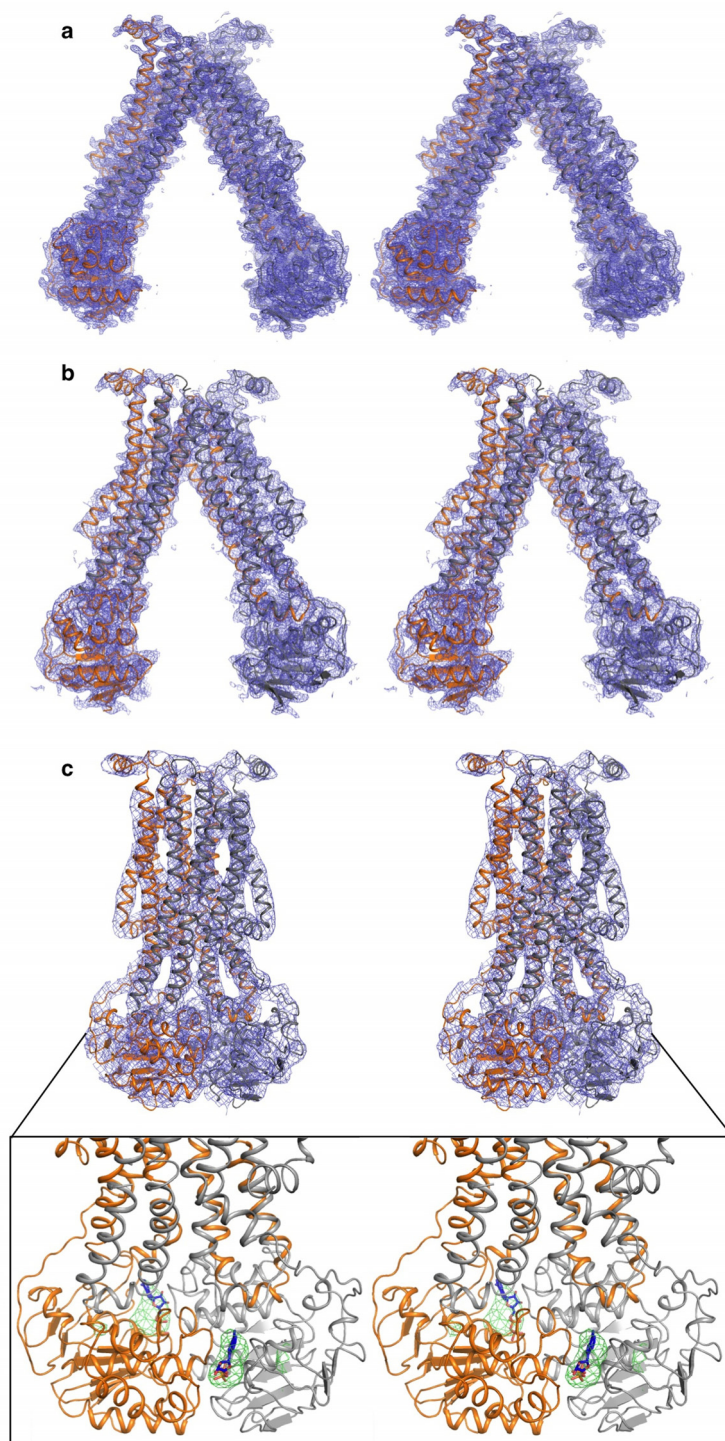
Extended Data Figure 4 | Electron density maps. a, b, Stereo views (cross-eyed) of the $2F_o - F_c$ electron density maps for the complete PglK dimer of the structures at 2.9 and 3.9 Å, respectively. **c**, Stereo view of the non-crystallographic symmetry (NCS)-averaged electron density map for the complete PglK dimer of the structure at 5.9 Å and close-up of the NBDs showing the $F_o - F_c$ electron density map for the bound ADP molecules. $2F_o - F_c$ and NCS maps are shown at 1.0σ level. $F_o - F_c$ maps are shown at 3.0σ level.
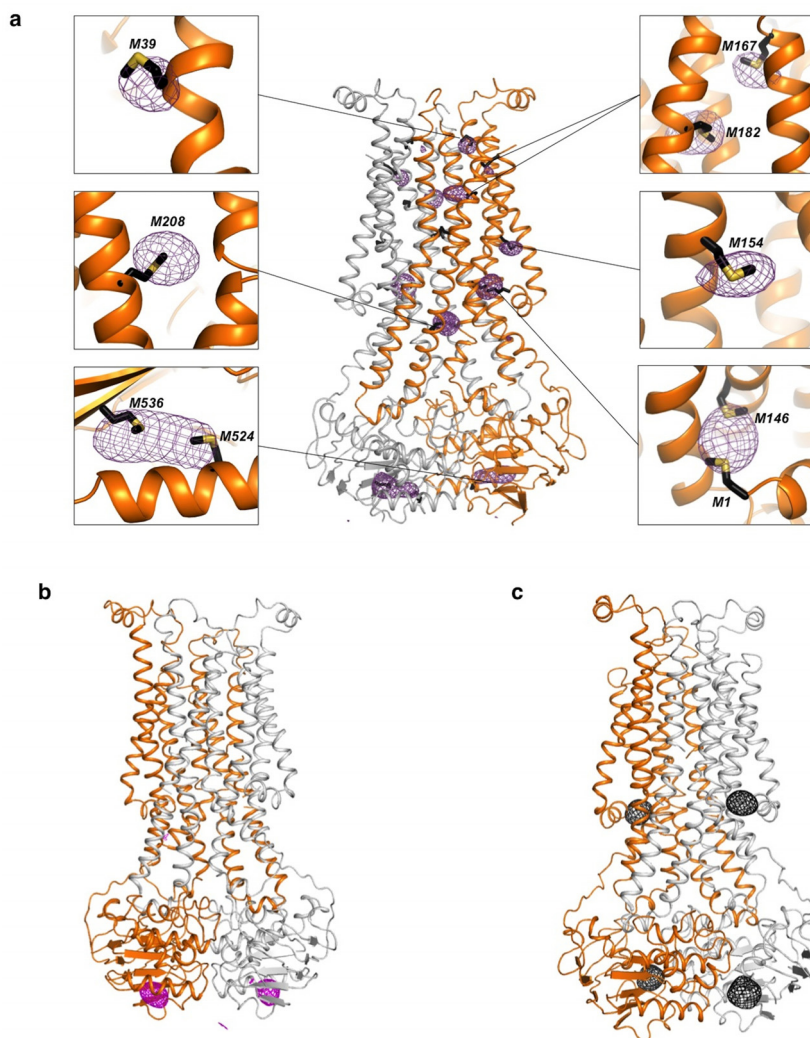
Extended Data Figure 5 | Validation of side-chain register of outward-occluded PglK model. a–c, Anomalous electron density maps define selenomethionine (**a**), cysteine-bound mercury (**b**) and PtCl$_4$ sites (**c**). Contour levels are between 4.0 to 5.0σ. In **a**, anomalous density was observed for 9 out of 10 selenomethionines of PglK.
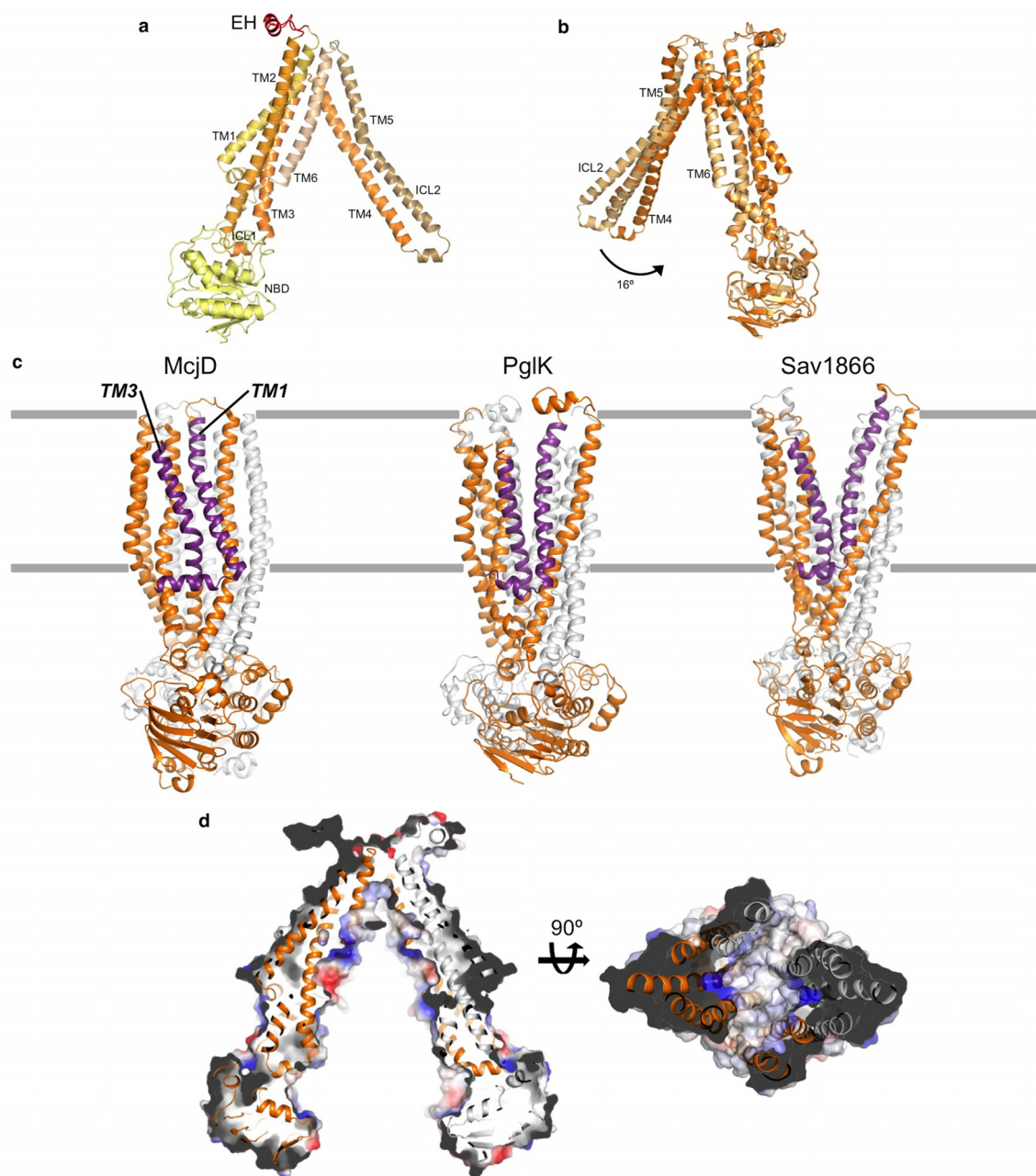
Extended Data Figure 6 | Structural features of PglK. a, Ribbon diagram of one PglK subunit depicting the secondary structure arrangement, based on the Sav1866 nomenclature[36]. **b,** Conformational changes of TM4 and TM5 visualized after superposition of subunits of apo-inward-1 (light orange) and apo-inward-2 (dark orange) structures. **c,** Structures of the antibacterial peptide ABC exporter McjD in occluded state (PDB code 4PLO), PglK in outward-occluded state, and the ABC exporter Sav1866 (PDB code 2HYD) in outward-open state. Transmembrane helices TM1 and TM3 (purple) define the extent of the external opening. Subunits in each dimer are shown in orange and grey. **d,** Side and cytoplasmic view cut-off of PglK apo-inward-1 structure and vacuum electrostatic surface representation showing the internal cavity.
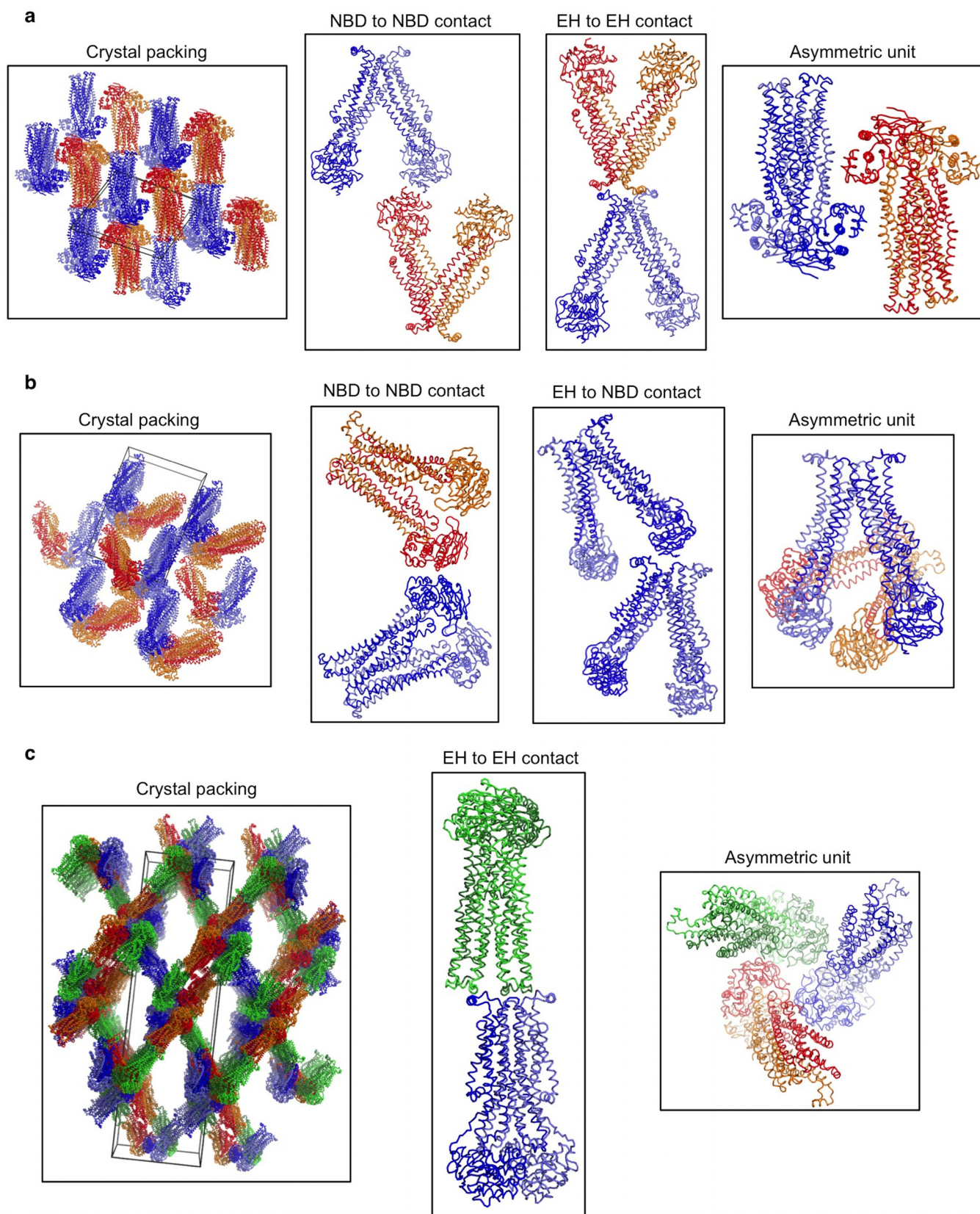
Extended Data Figure 7 | PglK crystal packing. a–c, The main crystal contacts of apo-inward-1 (2.9 Å resolution) (**a**), apo-inward-2 (3.9 Å resolution) (**b**), and outward-occluded (5.9 Å resolution) (**c**) states.
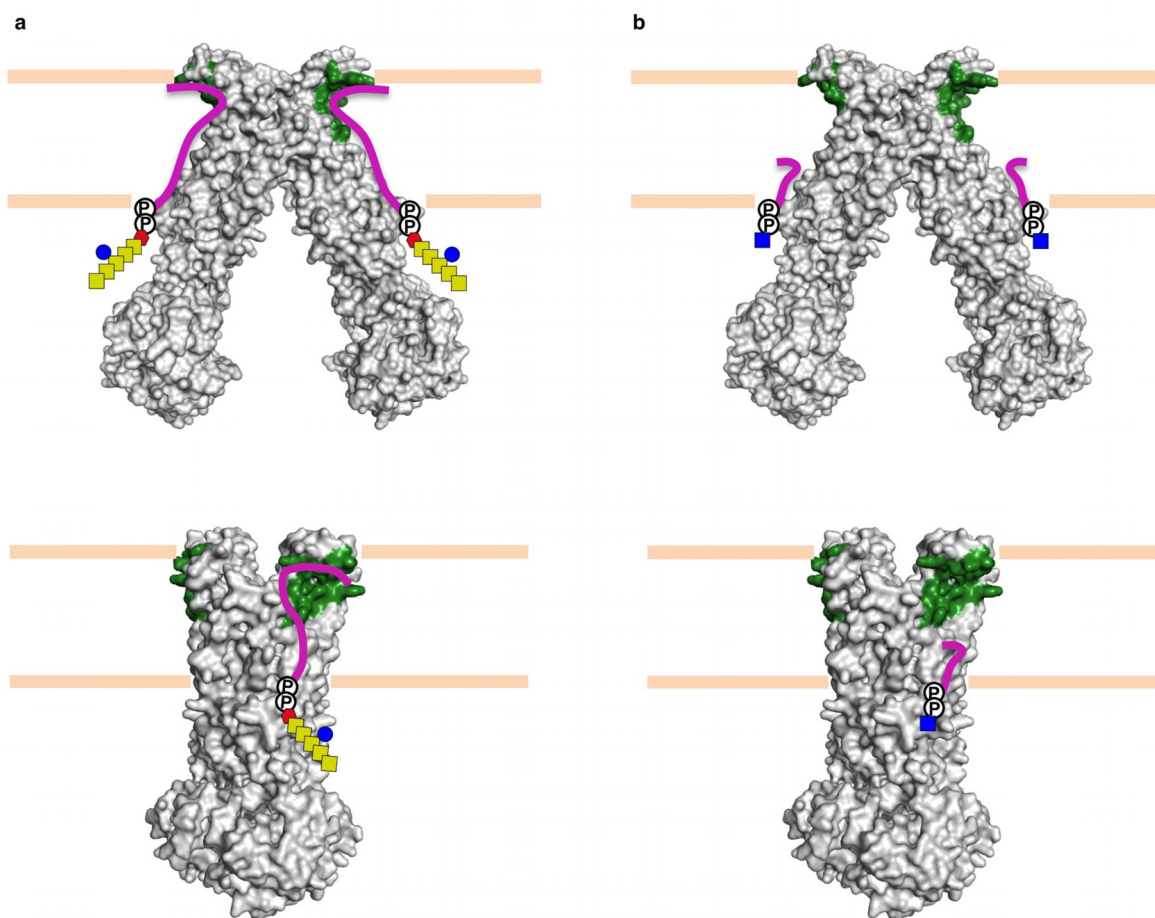
Extended Data Figure 8 | Putative interactions of native and synthetic LLO analogues with PglK in apo-inward and outward-occluded states. **a, b,** Native LLO (**a**) and synthetic LLO analogues (**b**). Red hexagon denotes di-*N*-acetylbacillosamine; yellow square denotes *N*-acetylgalactosamine; blue circle denotes glucose; and blue square denotes *N*-acetylglucosamine. The LLO polyprenyl tail is shown in purple.
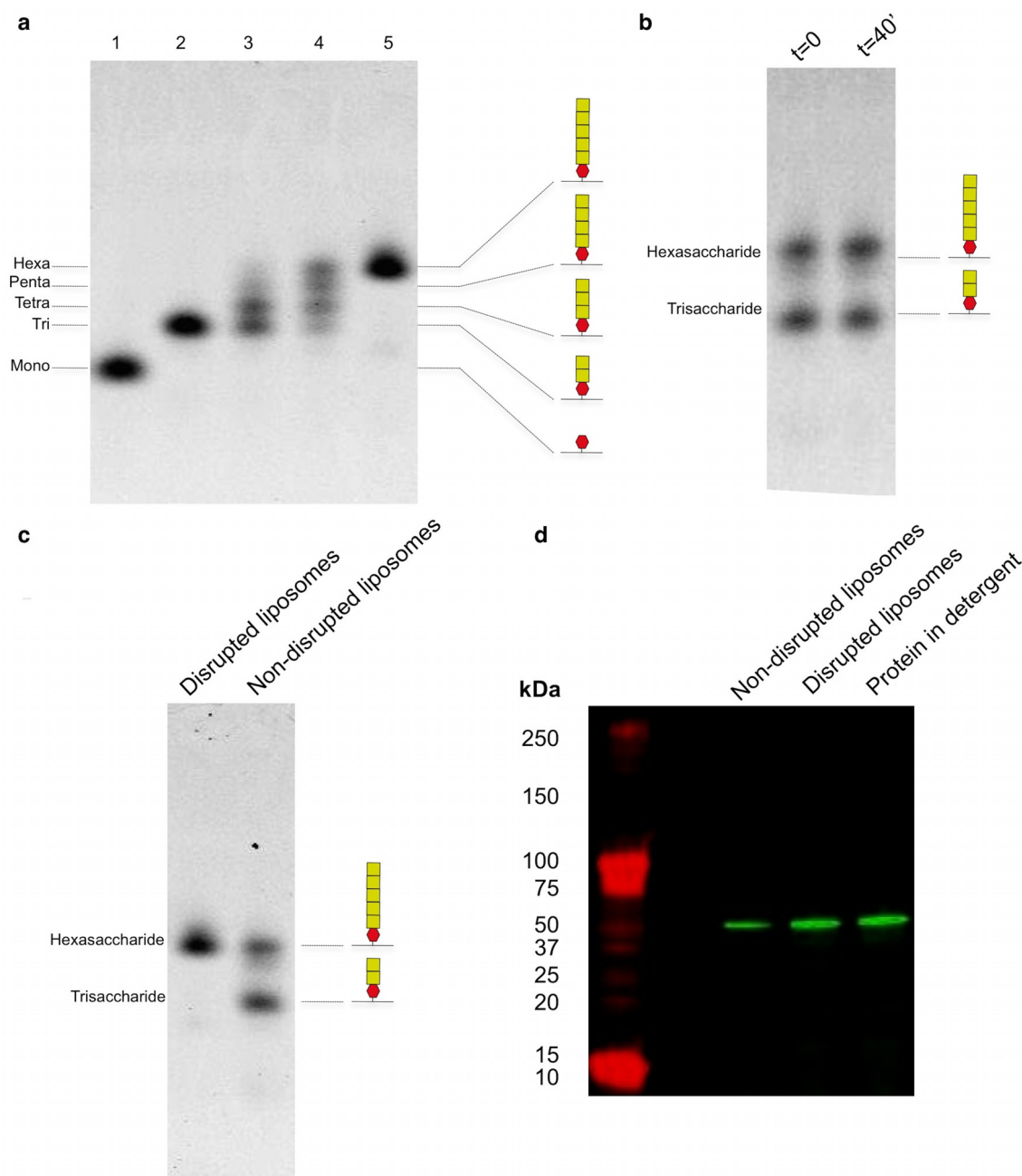
Extended Data Figure 9 | Product analysis and control experiments of PglK-catalysed *in vitro* flipping assay. a, Product analysis of tLLO glycosylation catalysed by PglH. The reaction products were analysed by *in vitro* glycosylation of a fluorescently labelled substrate peptide (DQNAT sequon) catalysed by PglB as reported earlier[30,31]. Depending on the presence and size of the *N*-glycan, peptides show a different mobility after Tricine–SDS–PAGE. Bands were visualized using a fluorescence gel scan (488 nm excitation and 526 nm emission). Lane 1, product of the deglycosylation reaction of hexasaccharide LLO catalysed by α-*N*-acetylgalactosaminidase, which removes terminal GalNAc molecules. This demonstrates the purity of the LLO used with respect to GalNAc. Lane 2, tLLO used in biochemical assays. Lane 3, products of tLLO glycosylation when the tLLO:GalNAc molar ratio is 2:1. Lane 4, products of tLLO glycosylation when the tLLO:GalNAc molar ratio is 1:1. Lane 5, products of tLLO glycosylation when the tLLO:GalNAc molar ratio is 1:10. **b**, PglK proteoliposomes incubated in the presence of 4 mM ADP. The level of tLLO in the external leaflet after *t* = 40 min remains unchanged

(104.2 ± 10.8%) relative to the amount at *t* = 0. ADP alone does not cause a decrease in the concentration of tLLO in the external liposomes leaflet, ruling out a potential sequestration of tLLO in the outward-occludded state of PglK. **c**, Determination of tLLO orientation in proteoliposomes. Disrupted liposomes (0.3% Triton X-100) and non-disrupted proteoliposomes were incubated with PglH in the presence of excess GalNAc. The amount of glycopeptide was determined from band intensities of fluorescence gel scans. 48.2 ± 7.5% of tLLO is located in the outer leaflet of the bilayer. **d**, Determination of PglK orientation in proteoliposomes. The fully functional mutant PglK(C269L/C352S/C386S/C549L/S544C) was reconstituted in proteoliposomes and labelled with negatively charged Alexa Fluor 488 C5 maleimide. The fluorescence of non-disrupted and disrupted proteoliposomes was compared. 51.8 ± 2.8% of the PglK molecules are oriented with NBDs facing outwards. Red hexagon denotes di-*N*-acetylbacillosamine; yellow square denotes *N*-acetylgalactosamine.

Extended Data Table 1 | X-ray data collection and refinement statistics

	Apo-inward-1 Native	Apo-inward-1 EMP	Apo-inward-2 Native	Out-occluded Native	Out-occluded SeMet	Out-occluded EMP	Out-occluded PtCl$_4^=$
Detergent	LMNG	LMNG	DDM	LMNG	LMNG	LMNG	LMNG
Data collection							
Wavelength (Å)	1.0000	1.0000	1.0000	1.0000	0.9779	1.0000	1.0722
Space group	P1	P1	P12$_1$1	P4$_3$2$_1$2	P4$_3$2$_1$2	P4$_3$2$_1$2	P4$_3$2$_1$2
Unit cell:							
a/b/c (Å)	96.6/114.4/145.6	95.6/115.2/148.0	93.4/183.3/136.3	200.0/200.0/693.9	200.0/200.0/694.0	202.9/202.9/685.7	201.9/201.9/692.2
α/β/γ (°)	68.4/73.3/68.7	67.8/73.9/69.1	90/106.7/90	90/90/90	90/90/90	90/90/90	90/90/90
Resolution (Å)	30-2.9	30-3.3	30-3.94	30-5.9	30-6.7	30-6.4	30-6.9
Completeness (%)	87.20[99.3]	95.6	86.20[99.6]	84.56 [98.3]	97.9	97.3	98.2
No. measured reflections	1176105 [1457567]	421629	480068 [804354]	1600808 [2330114]	1527912	1591766	1540412
No. unique reflections	206754 [258335]	206566	33489 [60617]	60330 [86104]	115835	118054	116442
I/σI	12.7(1.6) [10.4(0.6)]	6.55(1.55)	12.5(1.6) [11.1(0.3)]	18.1(2.0) [16(0.8)]	13.00(1.23)	8.45(1.13)	10.43(1.32)
R-merge (%)	11(140) [12(371)]		9.4(188) [11(158)]	20(257.4) [24(493)]			
CC$_{1/2}$ (%)	100(71) [100(51)]		100(70) [100(39)]	100(70) [100(42)]			
# of sites		12			54	6	12
Phasing power (iso./ano.)		0.557/0.788					
Figure of merit (FOM)		0.23/0.00*					
(acentric/centric)		0.86/0.00†					
Refinement							
R$_{work}$/R$_{free}$ (%)	23.15/26.71		27.4/32.9	28.4/31.7			
CC* (%)	100(90) [100(80)]		100(82) [100(73.2)]	100(92.4) [100(78)]			
R.m.s.d. Bonds (Å)	0.011		0.004	0.005			
R.m.s.d. Angles (°)	1.586		0.939	0.861			
Ramachandran plot:							
Outliers (%)	5.12		2.98	5.69			
Allowed (%)	10.50		12.01	11.83			
Favored (%)	84.39		85.01	82.47			
Average B-factor (Å^2)	71.6		166.3	335.5			
Completeness (%)	87.2		86.24	84.6			

Values in brackets are before anisotropic truncation; values in parentheses are for the last resolution shell. $R_{merge} = \Sigma_{hkl}\Sigma_i \left| I_i(hkl) - <I(hkl)> \right| / \Sigma_{hkl}\Sigma_i I_i(hkl)$

* Before density modification.

† After density modification.

The disruption of multiplanet systems through resonance with a binary orbit

Jihad R. Touma[1] & S. Sridhar[2]

Most exoplanetary systems in binary stars are of S-type, and consist of one or more planets orbiting a primary star with a wide binary stellar companion[1,2]. Planetary eccentricities and mutual inclinations can be large[3,4], perhaps forced gravitationally by the binary companion[4–6]. Earlier work on single planet systems[5,7–10] appealed to the Kozai–Lidov instability[11,12] wherein a sufficiently inclined binary orbit excites large-amplitude oscillations in the planet's eccentricity and inclination. The instability, however, can be quenched by many agents that induce fast orbital precession, including mutual gravitational forces in a multiplanet system[5,13]. Here we report that orbital precession, which inhibits Kozai–Lidov cycling in a multiplanet system, can become fast enough to resonate with the orbital motion of a distant binary companion. Resonant binary forcing results in dramatic outcomes ranging from the excitation of large planetary eccentricities and mutual inclinations to total disruption. Processes such as planetary migration[14,15] can bring an initially non-resonant system into resonance. As it does not require special physical or initial conditions, binary resonant driving is generic and may have altered the architecture of many multiplanet systems. It can also weaken the multiplanet occurrence rate in wide binaries, and affect planet formation in close binaries.

It is well known that planetary orbital precession periods are often many orders of magnitude longer than planetary orbital periods. But it has not been appreciated that the precession periods in S-type multiplanet systems can be comparable with (or even in resonance with) the orbital period of a distant binary companion. Even if the precessional periods are initially not exactly resonant with the binary period, they can be brought into resonance by various processes, one of which is planetesimal-driven planetary migration[14,16–18] (see Supplementary Information A). In the initial phase of migration a planet clears a gap in the planetesimal disk, resulting in either fast and long-range migration for a massive disk, or slow and short-range migration for a lighter disk[17]. Over the longer timescales of interest, mean motion resonances between the planet and planetesimals cause slow and long-range migration[15,16].

The fiducial system has two planets on initially coplanar orbits around a solar-mass primary star: an interior $10M_{\mathrm{Jup}}$ planet on a circular orbit with an initial semi-major axis of $a_{\mathrm{in}}^i = 5$ astronomical units (AU), and an exterior $10M_{\oplus}$ planet with initial semi-major axis a_{out}^i between 8 AU and 11 AU and an eccentricity $e_{\mathrm{out}}^i = 0.05$ (where M_{Jup} and $M_{\oplus}$ are the masses of Jupiter and Earth, respectively). The binary is also a solar-mass star with semi-major axis $a_{\mathrm{b}} > 100$ AU and corresponding period T_{b} and angular frequency n_{b}. The outer planet is allowed to migrate outward owing to interactions with a planetesimal disk, with its semi-major axis having a prescribed form whose characteristic timescale is τ. For the Solar System, there are plausible arguments that the migration time $\tau > 10^8$ yr (ref. 16) with a lower bound of $\tau > 10^7$ yr arguably needed to recover the properties of the Neptune Trojans[19]; we assume 2.5×10^7 yr $< \tau < 5 \times 10^8$ yr.

The physics of the problem appears clearest in a secular setting wherein planetary orbital motion is averaged over the fast orbital

phase, turning a point-mass planet into a shape-changing Gaussian wire[20]. We present a numerical simulation with an N-wire algorithm[21] and develop a mathematical model to understand the results. The binary orbit was chosen to be coplanar with the planetary system, and circular with $a_{\mathrm{b}} = 1{,}000$ AU, giving $T_{\mathrm{b}} = 22.36 \times 10^3$ yr (22.36 kyr). The fiducial two-planet system has two normal modes, the eccentric Laplace–Lagrange modes[20], which determine the rates at which the long axes of the nearly circular planetary orbits precess across the sky. With $a_{\mathrm{out}}^i = 10$ AU, the initial periods of these modes are 3.53×10^6 yr (3.53 Myr) (a slow mode associated with the massive inner planet) and 10.25 kyr (a fast mode reflecting the precession of the outer planet). Outward migration of the outer planet slows down the faster Laplace–Lagrange mode until its period approaches the binary period $T_{\mathrm{b}} = 22.36$ kyr. Figure 1a shows that the eccentricity of the outer planet $e_{\mathrm{out}} \approx 0.05$ until $a_{\mathrm{out}} \approx 11.89$ AU. Then the outer planet is captured into a resonance and e_{out} begins to increase, rising to 0.51 when $a_{\mathrm{out}} \approx 15$ AU at time $t \approx 4{,}055 \times T_{\mathrm{b}}$. Capture is also apparent in the behaviour of the resonant argument, $\phi_{\mathrm{res}}(t) = \varpi_{\mathrm{out}}(t) - n_{\mathrm{b}}t$, where ϖ_{out} is the apsidal longitude. From Fig. 1b we see that, after a period of circulation ϕ_{res} enters into libration at the resonance passage, with libration maintained for the full duration of the simulation. More details of the capture process are given in Supplementary Fig. 1a and b.

This capture phenomenon may be called the Laplace–Lagrange mode evection resonance (LLER) because it is closely related to the lunar evection resonance[22,23]. In Supplementary Information A we present an analytical model of LLER dynamics which is fourth order in planetary eccentricity and valid for arbitrary binary eccentricity. The theoretical prediction for the location of the exact resonance is shown as the dashed curve in Fig. 1a: whereas exact resonance is first met by a zero-eccentricity planet around $a_{\mathrm{out}} \approx 11.875$ AU, the planet circulating at $e_{\mathrm{out}} \approx 0.05$ encounters resonance a bit later (when $a_{\mathrm{out}} \approx 11.89$ AU), and then gets engulfed by a growing and migrating nonlinearly bounded resonance region. Prediction follows simulation in the
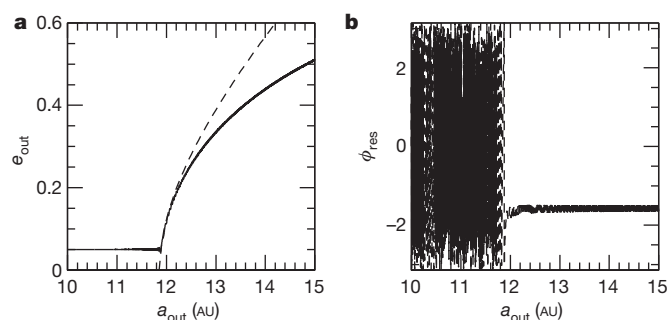

Figure 1 | Capture into the LLER. The fiducial N-wire experiment was performed with forced exponential migration[17], $a_{\mathrm{out}}(t) = a_{\mathrm{out}}^i \exp[t/\tau]$, with $a_{\mathrm{out}}^i = 10$ AU and $\tau = 10^4 T_{\mathrm{b}}$. **a**, Growth of e_{out} when it is captured in the migrating LLER. The dashed line is the prediction from the analytical fourth-order theory presented in Supplementary Information A. **b**, ϕ_{res} transitions from circulation to libration around 90° when captured in LLER.

[1]Department of Physics, American University of Beirut, PO Box 11–0236, Riad El-Solh, Beirut 1107 2020, Lebanon. [2]Raman Research Institute, Sadashivanagar, Bangalore 560 080, India.
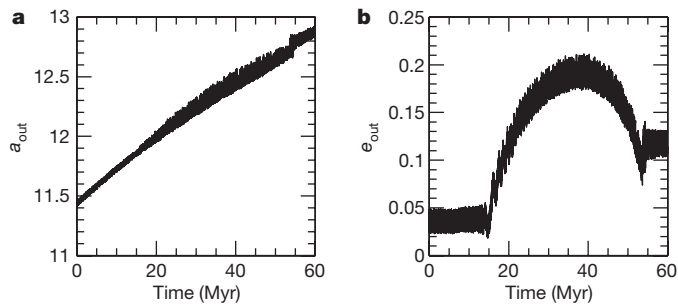
Figure 2 | LLER with PMMR for coplanar circular binary orbit. The first MERCURY experiment was conducted with damped migration[17], $a_{out}(t) = a_{fin} - \Delta a \exp[-t/\tau]$, with $a_{fin} = 14.5$ AU, $\Delta a = 3$ AU and $\tau = 4{,}500 T_b \approx 100$ Myr. **a**, Migration of a_{out} with signs of PMMR. **b**, LLER is encountered for $a_{out} \approx 11.88$ AU at $t \approx 15$ Myr, with resulting capture and initial growth of e_{out}, followed by near-reversal of that growth and ultimately ejection from LLER. The behaviour is typical of this migration scenario, with two-thirds of the systems (in a sample of a hundred, with $e^i_{out} = 0.05 \pm 0.02$, and randomly chosen outer-planet periapse) showing capture in the LLER, with variations in the final eccentricity resulting from multiple escape and recapture into the evection.

mean until $e_{out} \approx 0.2$; higher-order eccentricity expansion can improve the fit between model and simulation. Key dynamical features in and around resonance are discussed in Supplementary Information A.

Capture in LLER is probabilistic and depends on the strength of the resonance, the migration rate and the initial planetary eccentricity at which LLER is encountered. For $\tau \approx 10^4 T_b$ (or 224 Myr) and initial $e_{out} = 0.05$, N-wire simulations with ensembles of fiducial systems (a hundred such systems with randomly chosen outer-planet periapse) show capture of the planetary system in LLER in more than two-thirds of the cases; as discussed in Supplementary Information A, capture becomes certain as the migration rate is slowed down by two orders of magnitude. For $\tau \approx 4.5 \times 10^3 T_b$ (or 100 Myr), capture probability is smaller at about 0.3; for the shorter $\tau \approx 3 \times 10^3 T_b$, encounter of LLER with $e_{out} \approx 0.05$ leads to non-adiabatic passage, which could still excite e_{out} to 0.1. Capture is likely to improve in tighter S-type systems: for $a_b = 200$ AU and $a_{in} = 5$ AU, evection is crossed at $a_{out} \approx 7.43$ AU, with faster apse-precession and a stronger resonance, in the course of migration. Having described the broad secular skeleton of LLER, we note that the full problem is richer, owing to the interplay with planetary mean-motion resonances (PMMR). To study this, it is necessary to perform simulations that do not average over planetary mean motions. Below we present two such N-body simulations with the open-source package MERCURY[18,24].

The first MERCURY experiment is the unaveraged analogue of the N-wire simulation of Fig. 1, and its results for $\tau = 100$ Myr $\approx 4.5 \times 10^3 T_b$ are displayed in Fig. 2. Signs of PMMR are apparent in Fig. 2a, in the jumps experienced by the semi-major axis of the outer planet as it migrates. The system is captured in LLER with consequent growth of the eccentricity (Fig. 2b), and libration of the resonant argument (Supplementary Fig. 2a). What is remarkable though, and distinct from the N-wire experiment, is the non-monotonic behaviour of the mean eccentricity of the LLER-locked planet, leading to escape from LLER altogether, eventually settling at $e_{out} \approx 0.12$. We associate this behaviour with enhanced angular momentum exchange at PMMR, a novel variation on the classical evection that deserves further study. Of the four PMMR located near $a_{out} \approx 12.56$ AU the strongest is the 4:1 resonance, with argument $\phi_{4:1} = 4\lambda_{out} - \lambda_{in} - 3\varpi_{out}$; a short time segment of $\phi_{4:1}$ is shown in Supplementary Fig. 2b.

In the second MERCURY experiment, the binary orbit had eccentricity 0.4 and inclination 40°, which are modest values for wide binaries. Figure 3a shows, for $\tau = 4.4 \times 10^8$ yr $\approx 2 \times 10^4 T_b$, a 3:1 PMMR exciting e_{out} to 0.1 at $t \approx 21$ Myr, with capture in LLER at $t \approx 72$ Myr when $a_{out} \approx 11.92$ AU. As earlier, passage through the 4:1 resonance
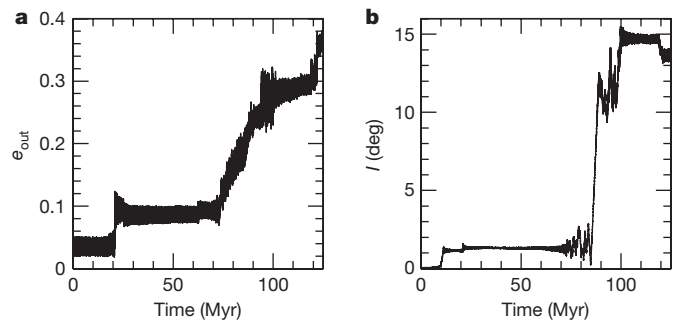

Figure 3 | LLER with PMMR for inclined and eccentric binary orbit. The second MERCURY experiment was conducted with exponential migration $a_{out}(t) = a^i_{out} \exp[t/\tau]$, with $a^i_{out} = 10$ AU and $\tau = 4.4 \times 10^8$ yr $= 2 \times 10^4 T_b$. **a**, The 3:1 PMMR around $t \approx 21$ Myr excites e_{out} to about 0.1. LLER is encountered around $a_{out} \approx 11.92$ AU, at $t \approx 72$ Myr. **b**, Mutual inclination I is first excited to 12° and then to 14.7°.

forces the outer planet out of LLER. This is followed by another excitation around 120 Myr, associated with passage through a 9:2 resonance; and then through a cluster of resonances around 14.63 AU. Meanwhile a_{out} grows with jumps at the PMMR (Supplementary Fig. 3a), and ϕ_{res} transits in and out of libration during LLER (Supplementary Fig. 3b). Both a_{out} and e_{out} diffuse until the planet is ejected from the system altogether. Ejection is not a necessary outcome, but is often associated with PMMR when both a_{out} and e_{out} are large. In Fig. 3b we follow the excitation of the mutual inclination to 12°, owing to coupling within LLER-lock, of eccentricity and inclination by a vertical resonance, which is followed by another excitation to 14.7°.

Of about 475 multiple planet systems in the Open Exoplanet Catalogue[25], about 124 (with a total of 322 planets) have reasonably well established planetary minimum masses and semi-major axes. Using minimum planetary masses, we computed Laplace–Lagrange mode periods for all 124 systems, and found a distribution peaked around log[mode period (yr)] ≈ 4 with full width at half maximum FWHM ≈ 2; for larger masses the distribution will shift towards shorter periods. The binary orbital period was sampled, for 500 yr to 10^5 yr, from an Öpik-type[26] distribution that is uniform in log[period (yr)], giving a mean of 3.85 and a full width of 2.3. The two distributions have broad overlap around a fiducial period of about 10^4 yr. Hence, even if a system is not initially in resonance, LLER is a distinct possibility because slow and long-range planetary migration[15,16] can bring it into resonance. This would be so especially in multiplanet systems, which have a richer spectrum of mode frequencies than do two-planet systems.

To go beyond parametrized migration models, in Supplementary Information B we present preliminary simulations of planetary migration using disks of planetesimals. Supplementary Fig. 6 shows that LLER is felt just as strongly by the migrating outer planet. Supplementary Information C explores the 'science-fiction' scenario of LLER in the outer Solar System were the Sun to have a distant companion star. Supplementary Fig. 7 shows that Saturn gets captured in LLER with high capture probabilities. LLER may be responsible for the recently reported[27] dearth of multiplanet systems in moderately wide ($a_b < 1{,}000$ AU) binaries. The extent to which LLER suppresses planet formation when $a_b < 20$ AU[28] needs to be assessed within planet formation studies[29]. The richer mode spectra and stronger PMMR in multi-mass systems open more pathways for LLER disruption, and could well relieve an initial multiplanet system of all but one of its planets.

1. Dvorak, R. Planetary orbits in double star systems. [in German] *Oesterreichische Akademie Wissenschaften Mathematisch naturwissenschaftliche Klasse Sitzungsberichte Abteilung* **191**, 423–437 (1982).

2. Raghavan, D. *et al.* A survey of stellar families: multiplicity of solar-type stars. *Astrophys. J.* **190** (Suppl.), 1–42 (2010).

3. Tamuz, O. *et al.* The CORALIE survey for southern extra-solar planets. XV. Discovery of two eccentric planets orbiting HD 4113 and HD 156846. *Astropart. Phys.* **480,** L33–L36 (2008).

4. Kaib, N. A., Raymond, S. N. & Duncan, M. J. 55 Cancri: a coplanar planetary system that is likely misaligned with its star. *Astrophys. J.* **742,** L24 (2011).

5. Holman, M., Touma, J. & Tremaine, S. Chaotic variations in the eccentricity of the planet orbiting 16 Cygni B. *Nature* **386,** 254–256 (1997).

6. Kaib, N. A., Raymond, S. N. & Duncan, M. Planetary system disruption by Galactic perturbations to wide binary stars. *Nature* **493,** 381–384 (2013).

7. Wu, Y. & Murray, N. Planet migration and binary companions: the case of HD 80606b. *Astrophys. J.* **589,** 605–614 (2003).

8. Fabrycky, D. & Tremaine, S. Shrinking binary and planetary orbits by Kozai cycles with tidal friction. *Astrophys. J.* **669,** 1298–1315 (2007).

9. Lithwick, Y. & Naoz, S. The eccentric Kozai mechanism for a test particle. *Astrophys. J.* **742,** 94 (2011).

10. Katz, B., Dong, S. & Malhotra, R. Long-term cycling of Kozai-Lidov cycles: extreme eccentricities and inclinations excited by a distant eccentric perturber. *Phys. Rev. Lett.* **107,** 181101 (2011).

11. Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67,** 591 (1962).

12. Lidov, M. L. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planet. Space Sci.* **9,** 719–759 (1962).

13. Takeda, G., Kita, R. & Rasio, F. A. Planetary systems in binaries. I. Dynamical classification. *Astrophys. J.* **683,** 1063–1075 (2008).

14. Malhotra, R. The origin of Pluto's peculiar orbit. *Nature* **365,** 819–821 (1993).

15. Murray, N., Hansen, B., Holman, M. & Tremaine, S. Migrating planets. *Science* **279,** 69–72 (1998).

16. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435,** 466–469 (2005).

17. Levison, H. F., Morbidelli, A., Gomes, R. & Backman, D. Planet migration in planetesimal disks. *Protostars Planets* **V,** 669–684 (2007).

18. Hahn, J. M. & Malhotra, R. Neptune's migration into a stirred-up Kuiper belt: a detailed comparison of simulations to observations. *Astron. J.* **130,** 2392–2414 (2005).

19. Lykawka, P. S., Horner, J., Jones, B. W. & Mukai, T. Origin and dynamical evolution of Neptune Trojans—I. Formation and planetary migration. *Mon. Not. R. Astron. Soc.* **398,** 1715–1729 (2009).

20. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* Ch. 7 (Cambridge Univ. Press, 1999).

21. Touma, J. R., Tremaine, S. & Kazandjian, M. V. Gauss's method for secular dynamics, softened. *Mon. Not. R. Astron. Soc.* **394,** 1085–1108 (2009).

22. Touma, J. & Wisdom, J. Resonances in the early evolution of the Earth–Moon system. *Astron. J.* **115,** 1653–1663 (1998).

23. Ćuk, M. & Stewart, S. T. Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning. *Science* **338,** 1047–1052 (2012).

24. Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304,** 793–799 (1999).

25. Rein, H. A proposal for community driven and decentralized astronomical databases and the Open Exoplanet Catalogue. Preprint at http://arxiv.org/abs/1211.7121 (2012).

26. Poveda, A., Allen, C. & Hernández-Alcántara, A. in *IAU Symposium* (eds Hartkopf, W. I., Harmanec, P. & Guinan, E. F.) Vol. 240, 417–425 (IAU, 2007).

27. Wang, J., Fischer, D. A., Xie, J.-W. & Ciardi, D. R. Influence of stellar multiplicity on planet formation. II. Planets are less common in multiple-star systems with separations smaller than 1500 AU. *Astrophys. J.* **791,** 111 (2014).

28. Wang, J., Xie, J.-W., Barclay, T. & Fischer, D. A. Influence of stellar multiplicity on planet formation. I. Evidence of suppressed planet formation due to stellar companions within 20 AU and validation of four planets from the Kepler multiple planet candidates. *Astrophys. J.* **783,** 4 (2014).

29. Rafikov, R. R. & Silsbee, K. Planet formation in stellar binaries. I. Planetesimal dynamics in massive protoplanetary disks. *Astrophys. J.* **798,** 69 (2015).

LETTER

Multi–gigaelectronvolt acceleration of positrons in a self-loaded plasma wakefield

S. Corde[1,2], E. Adli[1,3], J. M. Allen[1], W. An[4,5], C. I. Clarke[1], C. E. Clayton[4], J. P. Delahaye[1], J. Frederico[1], S. Gessner[1], S. Z. Green[1], M. J. Hogan[1], C. Joshi[4], N. Lipkowitz[1], M. Litos[1], W. Lu[6], K. A. Marsh[4], W. B. Mori[4,5], M. Schmeltz[1], N. Vafaei–Najafabadi[4], D. Walz[1], V. Yakimenko[1] & G. Yocky[1]

Electrical breakdown sets a limit on the kinetic energy that particles in a conventional radio-frequency accelerator can reach. New accelerator concepts must be developed to achieve higher energies and to make future particle colliders more compact and affordable. The plasma wakefield accelerator (PWFA) embodies one such concept, in which the electric field of a plasma wake excited by a bunch of charged particles (such as electrons) is used to accelerate a trailing bunch of particles. To apply plasma acceleration to electron–positron colliders, it is imperative that both the electrons and their antimatter counterpart, the positrons, are efficiently accelerated at high fields using plasmas[1]. Although substantial progress has recently been reported on high-field, high-efficiency acceleration of electrons in a PWFA powered by an electron bunch[2], such an electron-driven wake is unsuitable for the acceleration and focusing of a positron bunch. Here we demonstrate a new regime of PWFAs where particles in the front of a single positron bunch transfer their energy to a substantial number of those in the rear of the same bunch by exciting a wakefield in the plasma. In the process, the accelerating field is altered—'self-loaded'—so that about a billion positrons gain five gigaelectronvolts of energy with a narrow energy spread over a distance of just 1.3 metres. They extract about 30 per cent of the wake's energy and form a spectrally distinct bunch with a root-mean-square energy spread as low as

1.8 per cent. This ability to transfer energy efficiently from the front to the rear within a single positron bunch makes the PWFA scheme very attractive as an energy booster to an electron–positron collider.

Future high-energy particle colliders will operate at the frontier of particle physics, with particle energies in the range of several trillion electronvolts[3]. Beyond the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN), physicists envision building even bigger machines, such as the Future Circular Collider[4], which would collide protons at energies of around 100 TeV, but would require a tunnel of circumference approximately 100 km. Electron–positron linear colliders are also being considered, with proposed machines such as the International Linear Collider[5] (ILC) and the Compact Linear Collider[6] (CLIC). Based on existing radio-frequency technology, they are also expensive and tens of kilometres long. Looking beyond these machines, methods of building compact and efficient particle colliders—such as the muon collider[7], the laser wakefield accelerator[8] and the PWFA[9]—are under development. Of these, the PWFA has showed an energy gain of tens of billions of electronvolts in less than one metre[10], and has recently shown high-efficiency acceleration of an electron bunch with a narrow energy spread containing a substantial charge, at a high energy gain per unit length (or gradient) (ref. 2). However, for a future PWFA-based
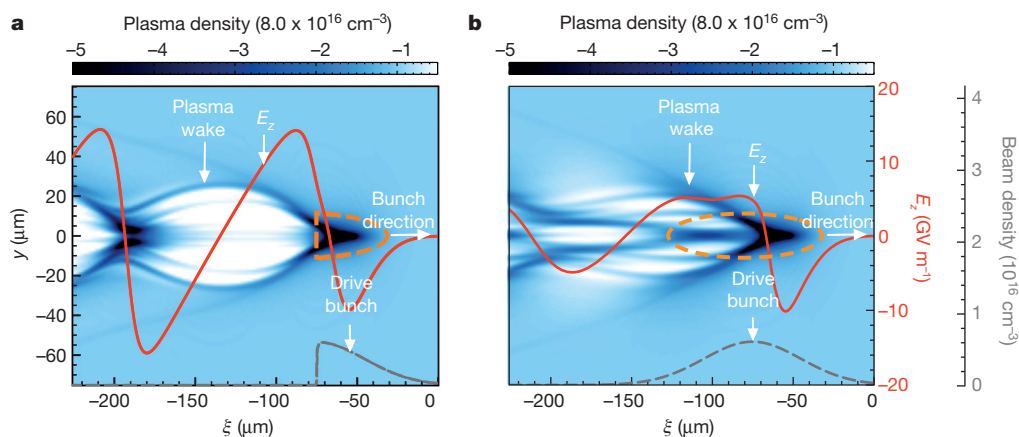

Figure 1 | Simulated plasma wakes driven by short and intense positron bunches. The electron plasma density in the y–ξ plane is shown after a propagation distance of $s = ct = 135$ cm into the plasma, where y is the dimension transverse to the direction of motion of the bunch, and $\xi = z - ct$ is the dimension parallel to the motion. In both panels the one-third-of-the-maximum contour of the initial positron bunch density is represented by the orange dashed line, the on-axis density profile of the initial positron bunch by the grey dashed line, and the on-axis longitudinal electric field E_z by the red

solid line. The beam and plasma parameters (described in the Fig. 3 legend) are the same as those in the experiment. **a,** The unloaded plasma wake, for which no positrons are being accelerated by the wake because the bunch has been terminated just as E_z reverses sign. **b,** The self-loaded plasma wake, where the trailing particles of a single bunch extract energy from the wake excited by the particles in the front. In **a** and **b**, the colour scale represents the perturbed plasma electron density (the plasma density prior to the passage of the beam is constant at 8×10^{16} cm^{-3}).

[1]SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [2]LOA, ENSTA ParisTech, CNRS, Ecole Polytechnique, Université Paris-Saclay, 91762 Palaiseau, France. [3]Department of Physics, University of Oslo, 0316 Oslo, Norway. [4]Department of Electrical Engineering, University of California Los Angeles, Los Angeles, California 90095, USA. [5]Department of Physics and Astronomy, University of California Los Angeles, Los Angeles, California 90095, USA. [6]Department of Engineering Physics, Tsinghua University, Beijing 100084, China.

particle collider, it is imperative to demonstrate that the antimatter counterpart of the electron, the positron, can also be accelerated in a PWFA at high gradient and with high-energy efficiency.

The longitudinal component of the electric field E_z associated with a wake produced by the passage of either an intense ultra-relativistic electron or positron bunch through a plasma can, in principle, be used to accelerate positrons. In both cases, a dense ($n_b > n_p$), tightly focused ($k_p\sigma_r < 1$), short ($k_p\sigma_z < 1$), ultra-relativistic ($\gamma \gg 1$) drive bunch (electron or positron) can be used to excite a nonlinear (non-sinusoidal) wake in a plasma. Here n_b, n_p, k_p, σ_r, σ_z and γ are the bunch density, the plasma density, the wavenumber of the plasma wave, the root-mean-square (r.m.s.) focused transverse spot size, the r.m.s. longitudinal size and the Lorentz factor of the bunch, respectively. However, the nonlinear wakes produced by the two types of drivers are qualitatively different[1]. With an electron driver, a region devoid of plasma electrons, called an ion cavity, is formed as these electrons are blown out by the transverse electric field of the bunch[11,12]. Within this ion cavity, the transverse force is defocusing for positrons and so prevents positron acceleration. To avoid this problem, the use of a hollow plasma channel to produce wakes without a focusing force[13–15], or the use of Laguerre–Gaussian laser pulses to drive doughnut-shaped wakes with a strong focusing force for positrons[16], have been suggested.

In contrast to the electron-driven wake, when an otherwise similar positron bunch is used, plasma electrons that are radially located within a few plasma skin depths (k_p^{-1}, it the penetration depth of a low-frequency electromagnetic wave in a plasma) from the bunch are attracted inward (rather than being expelled) by the transverse electric field of the bunch[17–19]. As the plasma electrons flow in, positrons experience a negative or decelerating E_z, losing energy as they do work on the plasma electrons. Once most plasma electrons have crossed the propagation axis, E_z abruptly switches sign from negative to positive and becomes accelerating for positrons. If there are no positrons to sample the accelerating field, the wake is said to be unloaded because no energy is extracted from it. This is shown in Fig. 1a, taken from a simulation using the three-dimensional particle-in-cell code QuickPIC[20,21]. In this case, plasma electrons flow outward after they cross the propagation axis, which leads to the formation of a cavity in which the ion density exceeds the electron density. When a sufficient number of positrons are sampling the accelerating field, a large number of electrons crossing the axis remain close to the axis (see Fig. 1b). Consequently the longitudinal and the transverse fields are both strongly altered, that is, loaded. Owing to the presence of the plasma electrons on axis, the accelerated positrons are guided along the length of the plasma. These positrons can extract a substantial fraction of the wake's energy and alter the shape of the E_z field, which becomes more uniform and produces a narrow-energy-spread peak in the accelerated part of the positron spectrum. This process occurs in the positron-driven wake without the need for a distinct trailing bunch[22]. In other words, the front of a single positron bunch can excite a wake in a plasma while the rear of the same bunch loads and extracts energy from this wake (as in Fig. 1b). This regime is referred to as the self-loaded plasma wakefield.

The formation of an accelerated narrow-energy-spread positron bunch was discovered in an experiment conducted at the SLAC's Facility for Advanced Accelerator Experimental Tests[23] (FACET), using its 20.35-GeV positron beam (see Methods). A single bunch containing approximately 1.4×10^{10} positrons and having an r.m.s. bunch length in the range 30–50 μm was focused to an r.m.s. transverse spot size of less than 100 μm at the entrance of a lithium plasma (vacuum beam density $n_b \approx 0.2$–1×10^{16} cm^{-3}). The plasma is produced by laser ionization of a 1.15-m-long lithium vapour column of uniform density that has 15-cm-long density up- and down-ramps on either end[24,25] (see Methods). The electron density of the plasma was set to $n_p = 8 \times 10^{16}$ cm^{-3} by controlling the pressure and the temperature of the lithium vapour. After the interaction with the plasma, an imaging spectrometer consisting of a quadrupole magnet doublet, a

strong dipole magnet and a Cherenkov detector[26] was used to characterize the energy spectrum of the positron beam (see Methods). The quadrupole magnet doublet was set up to image where the positron beam exits the plasma into the plane of the detector for a given positron energy: the energy set-point E_{image}. To record both the decelerated and the accelerated parts of the positron spectrum, E_{image} of the quadrupole magnet doublet was varied in increments of 2.5 GeV from 10.35 GeV to 27.85 GeV.

Figure 2a and b show the accelerated part of the final energy spectrum of the positron bunch after its passage through the plasma, where E_{image} is set to 22.85 GeV and 25.35 GeV, respectively. The accelerated positrons in Fig. 2a and b have peaks at 24.75 ± 0.27 GeV and 26.11 ± 0.35 GeV. These two examples show that the accelerated positrons have a narrow energy peak, albeit on top of a broader 'shoulder'. The spectral peaks are fitted by an asymmetric Gaussian function. The r.m.s. energy spreads associated with the fit of the peaks (red dashed lines) are 1.8% and 2.2% r.m.s., respectively, while the full width at half-maximum (FWHM) energy spreads associated with the experimental spectra (black solid lines) are 4.0% and 6.1%, respectively.
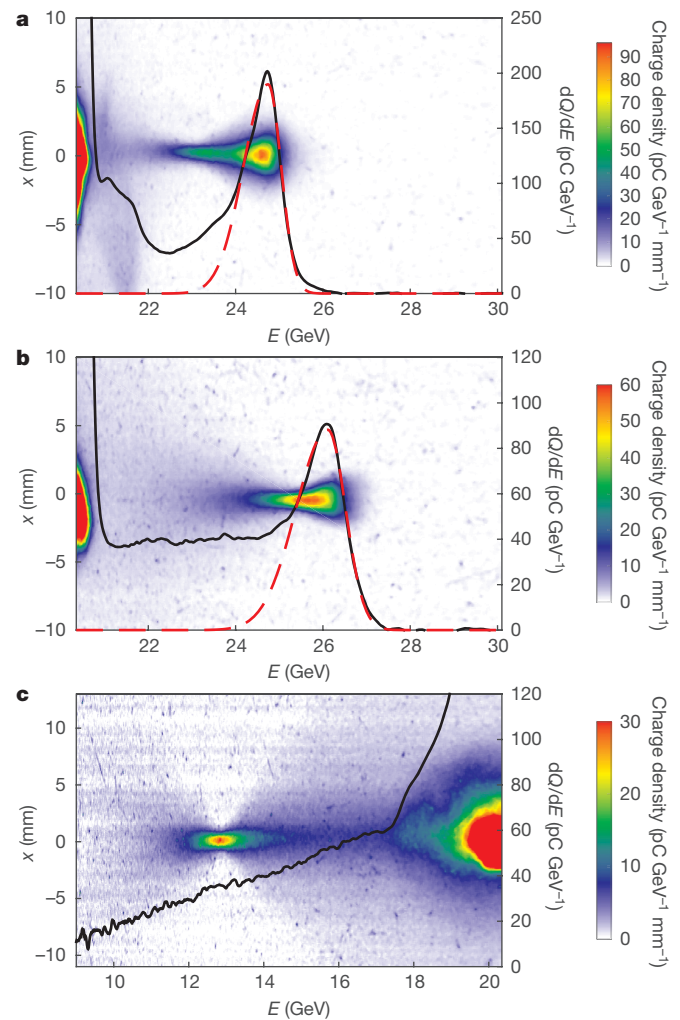

Figure 2 | Experimental positron energy spectra. **a**, **b**, Two examples of the accelerated portion of the spectrum of the positrons after the interaction with the plasma with E_{image} at 22.85 GeV (**a**) and 25.35 GeV (**b**). **c**, An example of the decelerated portion of the positron spectrum with E_{image} at 12.85 GeV. In all panels, the density of charge per unit energy and length of the dispersed positron beam profile is shown in colour, and the spectral charge density dQ/dE is represented by the black solid line (right scale). Asymmetric Gaussian fits to the peaks in **a** and **b** are shown as red dashed lines. Particles that do not participate in the interaction appear to be saturated at the initial beam energy, 20.35 GeV.
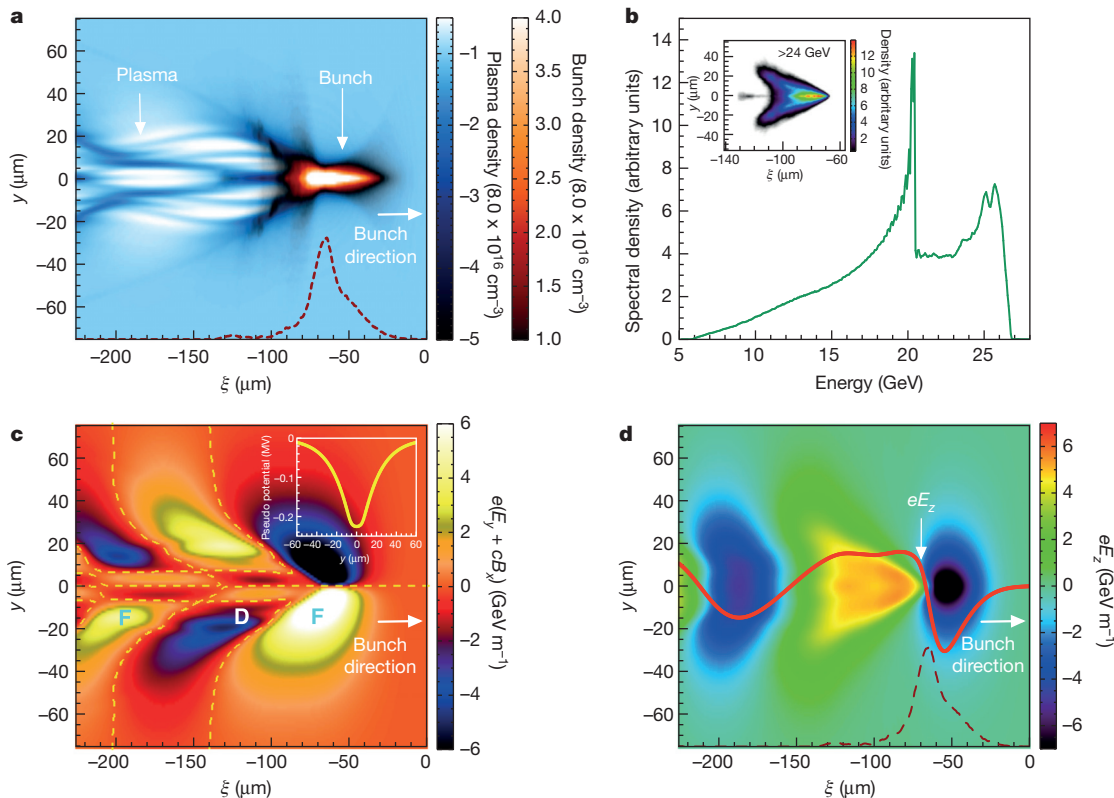
Figure 3 | Particle-in-cell simulation. a, Electron plasma density and positron bunch density in the y–ξ plane. **b,** Final spectrum of the positron bunch after its interaction with the plasma. The inset shows the projected density in the y–ξ plane of the positrons with energies greater than 24 GeV. **c,** The transverse force along y experienced by the positrons is $F_y \approx e(E_y + cB_x)$ (where e is the positron elementary charge, E_y is the electric field along y and B_x is the magnetic field along x). The dashed lines represent the $F_y = 0$ contour. F and D indicate, respectively, the focusing and defocusing regions of the wake. The inset shows the pseudo potential $\psi(x = 0, y, \xi = \xi_0)$ (ψ being given implicitly by $\boldsymbol{F}_\perp = -\nabla_\perp e\psi$, where $\boldsymbol{F}_\perp$ is the transverse force experienced by the positrons) along a line taken at $\xi_0 = -75\,\mu m$. **d,** Longitudinal electric field E_z. In **a** and **d**, the on-axis density profile of the positron bunch is represented by the red dashed line. In this simulation, the bunch contains 1.4×10^{10} positrons and has transverse and longitudinal spot sizes of $\sigma_r = 70\,\mu m$ and $\sigma_z = 30\,\mu m$. The bunch energy is 20.35 GeV and the plasma density is $8 \times 10^{16}\,cm^{-3}$. The plasma wake and bunch properties shown in **a**–**d** are taken after a propagation distance of $s = ct = 135\,cm$.

The positron charge contained in the peak (under the red dashed curve) is 207 ± 14 pC for the shot of Fig. 2a and 126 ± 8 pC for the shot of Fig. 2b. We have found experimentally that the energy gain of this peak can vary shot to shot from 3 GeV to 10 GeV.

The contrast between the peak and the shoulder also varied shot to shot, and in some cases, the accelerated part of the spectrum was flat or decreasing (see Extended Data Fig. 1). The narrowest peaks were observed when the maximum energy gain was about 5 GeV. For an acceleration length of 1.3 m, this energy gain corresponds to an average loaded gradient of about 3.8 GeV m^{-1}. This lower-energy shoulder can be suppressed by a proper collimation scheme using a series of magnets. Simulations discussed in Methods show that the peak energy and the peak-to-shoulder ratio are sensitive to σ_z, the details of the longitudinal shape of the bunch (that is, the rise time versus the fall time), σ_r, and the normalized transverse emittance of the bunch (see Extended Data Fig. 2). In the experiment, several of these parameters fluctuated from one shot to the next, and it was not possible to determine which parameter affected the experimental outcome the most. Nevertheless, these results demonstrate the self-formation of a narrow-energy-spread positron bunch with multi-gigaelectronvolt energy gain in a metre-scale plasma accelerator.

Figure 2c shows the decelerated part of the positron spectrum, with E_{image} set to 12.85 GeV. The spectral charge density is continuous and decreases monotonically from 20.35 GeV to less than 10 GeV. The amount of energy that is transferred to the wake is computed as the total amount of energy lost by the decelerated positrons. For the shot of Fig. 2c, a decelerated charge of 511 ± 34 pC has transferred 2.40 J of energy to the plasma wake. The energy extracted from the wake by the accelerated charge in the narrow peaks of Fig. 2a and b was calculated to be 0.86 J and 0.69 J, respectively. The energy extraction efficiency is obtained by taking the ratio of the energy gained by the accelerated charge contained in the peak to the energy transferred to the wake by the decelerated charge. Because the spectral charge density is only accurately measured for energies near E_{image}, the amount of energy lost by the decelerated charge cannot be determined for the shots shown in Fig. 2a and b. The energy extraction efficiency and its error are thus estimated using the average amount of energy loss and its shot-to-shot variation, which are obtained using data acquired with E_{image} ranging from 10.35 GeV to 20.35 GeV (see Methods). The resultant average energy loss is found to be 2.54 J, with variations of ± 0.3 J from shot to shot. The fraction of energy extracted from the wake by the charge in the peaks of Fig. 2a and b is thus estimated to be $34\% \pm 5\%$ and $27\% \pm 4\%$, respectively.

To interpret these experimental results, three-dimensional particle-in-cell simulations for beam and plasma parameters similar to those used in the experiment are performed with the code QuickPIC[20,21]. In particular, the bunch has asymmetric normalized emittances in the two transverse dimensions of 200 mm mrad (in x) and 50 mm mrad (in y). Results are presented in Fig. 3. During the first 10 cm of propagation in the flat-density region of the plasma, feedback of the forces exerted by the plasma on the positron bunch and vice versa causes the positron bunch and the plasma wake to evolve rapidly, until they reach a quasi-steady state. In this quasi-steady state, the longitudinal and transverse fields vary slowly in time (see Supplementary Video 1).

Figure 3a shows that after 135 cm of propagation through the plasma the initially Gaussian positron bunch evolves to form an arrowhead-shaped structure. Interestingly, a small number of positrons located in the region $-140\,\mu m < \xi < -90\,\mu m$ (where $\xi = z - ct$ is the dimension parallel to the motion, comoving at the speed of light) are focused by the plasma electrons close to the bunch axis, as seen in the on-axis positron bunch density profile of Fig. 3a. Figure 3b shows the energy spectrum of the positrons at this point in the plasma. It shows that the initial 20.35-GeV positron bunch has both lost and gained energy. The energy loss monotonically extends down to 7 GeV but the energy gain feature shows a clear high-energy peak centred around 25 GeV, on top of a broader energy shoulder. The inset in Fig. 3b shows that most positrons with energies exceeding 24 GeV reside close to the beam axis in the region $-100\,\mu m < \xi < -70\,\mu m$. The transverse force (see Fig. 3c and Extended Data Fig. 3) confines positrons close to the axis in a pseudo potential well, as shown in the inset of Fig. 3c.

This large positron charge experiencing an accelerating electric field can load and thereby extract a substantial fraction of the energy stored in the wake. This self-loading of the wake can result in a nearly uniform longitudinal electric field along the length of the accelerated charge. These positrons thus gain approximately the same energy, which leads to the formation of a narrow-energy-spread accelerated peak in the final positron spectrum (see Fig. 3b). The colour plot of the E_z field and its on-axis lineout in Fig. 3d clearly show that the peak decelerating field is larger than the peak accelerating field. Further, the volume with the highest decelerating field is small compared to the extent of the bunch. This leads to a final positron energy spectrum that has a long energy loss tail (see Fig. 3b), in good agreement with the experimental results shown in Fig. 2.

In conclusion, we have discovered a mechanism that accelerates positrons with fields of several gigaelectronvolts per metre in a plasma. Using a nonlinear plasma wake driven by a single positron bunch, a substantial number of positrons is shown to be accelerated and guided over a metre-scale plasma. The accelerating electric fields exceed by two orders of magnitude those available with radio-frequency technology. The large number of accelerated positrons loads the wake and extracts almost a third of its energy. This self-loading flattens the longitudinal electric field of the wake, resulting in a narrow-energy-spread spectrum. These results show that the PWFA concept may be a solution for an energy-boosting upgrade to a linear collider[27].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Joshi, C. & Katsouleas, T. Plasma accelerators at the energy frontier and on tabletops. *Phys. Today* **56**, 47–53 (2003).
2. Litos, M. *et al.* High-efficiency acceleration of an electron beam in a plasma wakefield accelerator. *Nature* **515**, 92–95 (2014).
3. The Particle Physics Project Prioritization Panel (P5). Subpanel of the High Energy Physics Advisory Panel (HEPAP). *Building for Discovery, Strategic Plan for U.S. Particle Physics in the Global Context.* http://www.usparticlephysics.org/p5/ (2014).
4. Zimmermann, F., Benedikt, M., Schulte, D. & Wenninger, J. Challenges for highest energy circular colliders. In *Proc. IPAC2014* MOXAA01, http://accelconf.web.cern.ch/accelconf/IPAC2014/papers/moxaa01.pdf (JACoW, 2014).
5. Barish, B. & Yamada, S. (eds) *The International Linear Collider Technical Design Report.* ILC-REPORT-2013–040, http://www.linearcollider.org/ILC/Publications/Technical-Design-Report (ILC, 2013).
6. Aicheler, M., *et al. A Multi-TeV Linear Collider based on CLIC Technology: CLIC Conceptual Design Report.* CERN-2012–007, http://dx.doi.org/10.5170/CERN-2012-007 (CERN, 2012).
7. Alexahin, Y. *et al.* Muon Collider Higgs Factory for Snowmass 2013. In *Proc. Community Summer Study 2013: Snowmass on the Mississippi.* Preprint at http://arxiv.org/abs/1308.2143 (2013).
8. Schroeder, C. B., Esarey, E., Geddes, C. G. R., Benedetti, C. & Leemans, W. P. Physics considerations for laser-plasma linear colliders. *Phys. Rev. ST Accel. Beams* **13**, 101301 (2010).
9. Adli, E. *et al.* A beam driven plasma-wakefield linear collider: from Higgs factory to multi-TeV. In *Proc. Community Summer Study 2013: Snowmass on the Mississippi.* Preprint at http://arxiv.org/abs/1308.1145 (2013).
10. Blumenfeld, I. *et al.* Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator. *Nature* **445**, 741–744 (2007).
11. Rosenzweig, J. B., Breizman, B., Katsouleas, T. & Su, J. J. Acceleration and focusing of electrons in two-dimensional nonlinear plasma wake fields. *Phys. Rev. A* **44**, R6189–R6192 (1991).
12. Lu, W., Huang, C., Zhou, M., Mori, W. B. & Katsouleas, T. Nonlinear theory for relativistic plasma wakefields in the blowout regime. *Phys. Rev. Lett.* **96**, 165002 (2006).
13. Lee, S., Katsouleas, T., Hemker, R. G., Dodd, E. S. & Mori, W. B. Plasma-wakefield acceleration of a positron beam. *Phys. Rev. E* **64**, 045501(R) (2001).
14. Kimura, W. D., Milchberg, H. M., Muggli, P., Li, X. & Mori, W. B. Hollow plasma channel for positron plasma wakefield acceleration. *Phys. Rev. ST Accel. Beams* **14**, 041301 (2011).
15. Yi, L. Q. *et al.* Positron acceleration in a hollow plasma channel up to TeV regime. *Sci. Rep.* **4**, 4171 (2014).
16. Vieira, J. & Mendonça, J. T. Nonlinear laser driven donut wakefields for positron and electron acceleration. *Phys. Rev. Lett.* **112**, 215001 (2014).
17. Hogan, M. J. *et al.* Ultrarelativistic-positron-beam transport through meter-scale plasmas. *Phys. Rev. Lett.* **90**, 205002 (2003).
18. Blue, B. E. *et al.* Plasma-wakefield acceleration of an intense positron beam. *Phys. Rev. Lett.* **90**, 214801 (2003).
19. Muggli, P. *et al.* Halo formation and emittance growth of positron beams in plasmas. *Phys. Rev. Lett.* **101**, 055001 (2008).
20. Huang, C. *et al.* QuickPIC: a highly efficient fully parallelized PIC code for plasma-based acceleration. *J. Phys. Conf. Ser.* **46**, 190–199 (2006).
21. An, W., Decyk, V. K., Mori, W. B. & Antonsen, T. M. Jr. An improved iteration loop for the three dimensional quasi-static particle-in-cell algorithm: QuickPIC. *J. Comput. Phys.* **250**, 165–177 (2013).
22. Tzoufras, M. *et al.* Beam loading in the nonlinear regime of plasma-based acceleration. *Phys. Rev. Lett.* **101**, 145002 (2008).
23. Hogan, M. J. *et al.* Plasma wakefield acceleration experiments at FACET. *New J. Phys.* **12**, 055030 (2010).
24. Muggli, P. *et al.* Photo-ionized lithium source for plasma accelerator applications. *IEEE Trans. Plasma Sci.* **27**, 791–799 (1999).
25. Green, S. Z. *et al.* Laser ionized preformed plasma at FACET. *Plasma Phys. Contr. Fusion* **56**, 084017 (2014).
26. Adli, E., Gessner, S. J., Corde, S., Hogan, M. J. & Bjerke, H. H. Cherenkov light-based beam profiling for ultrarelativistic electron beams. *Nucl. Instrum. Methods Phys. Res. A* **783**, 35–42 (2015).
27. Lee, S. *et al.* Energy doubler for a linear collider. *Phys. Rev. ST Accel. Beams* **5**, 011001 (2002).

Author Contributions All authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.C. (corde@slac.stanford.edu).

METHODS

Positron beam. In the experiment reported here we used the 20.35-GeV, ~2.2-nC positron bunches delivered by the FACET facility to the experimental area situated in Sector 20 (see Extended Data Fig. 4). The positrons at FACET originate from the electromagnetic shower produced when a 20.35-GeV electron beam is sent into a six-radiation-length tungsten alloy target[28]. An electron beam is therefore first accelerated in the FACET 2-km-long linear accelerator, before being sent to the positron target in Sector 19. A large number of electron–positron pairs are created during the electromagnetic cascade in the target. Positrons exiting the target are captured and transported by the Positron Return Line to the start of the linear accelerator, where they are accelerated and injected into the South Damping Ring of Sector 2. After cooling the positron beam in the damping ring, the beam is accelerated in the linear accelerator and undergoes three stages of compression[23], (1) in the ring-to-linear accelerator section, (2) in Sector 10 and (3) in Sector 20. At the exit of the Sector 20 bunch compressor, the beam has a fully compressed bunch length in the range of 30–50 μm (r.m.s.), corresponding to a duration of 100–170 fs (r.m.s.). Five quadrupole magnets (the final focus system) focus the beam to a spot size of less than 100 μm (r.m.s.) in both transverse dimensions at the entrance of the plasma. The final focus is configured so as to have beta functions (the propagation distance from the waist at which the beam size is increased by $2^{1/2}$) of 50 cm in x and 500 cm in y at focus.

Plasma source. A column of lithium vapour is confined in a heat-pipe oven by a helium buffer gas at both ends[24]. The lithium atomic density is uniform over a length of 1.15 m, and has 15-cm-long ramps on either end, giving a FWHM length of 1.3 m. The lithium atomic density in the plateau region is set to 8×10^{16} cm^{-3} by controlling the pressure and the temperature in the heat-pipe oven. The lithium vapour is ionized before the passage of the positron bunch by a Ti:sapphire laser pulse[25]. The laser has an energy of about 20 mJ, a FWHM pulse duration of 200 fs and is focused by a 0.75° Axicon lens, which produces a zero-order Bessel intensity profile over the full length of the column of lithium vapour. The laser focus line is aligned to the positron bunch axis over the entire length of the oven.

Energy spectrum and efficiency. The energy of the positrons is characterized using a Cherenkov imaging spectrometer[26]. A dipole magnet, whose equivalent length and magnetic field are 97.8 cm and 0.8 T, respectively, deflects the positrons vertically. Positrons with an energy of 20.35 GeV experience a 11.46 mrad deflection. At the location of the detector, the vertical position of a positron provides a measurement of its energy. A quadrupole magnet doublet is used to image where the positron beam exits the plasma into the plane of the detector. The imaging conditions hold only for a given positron energy E_{image}. The dispersion from the dipole magnet was high enough so that positrons with energies in a band of ~3 GeV around E_{image} are accurately resolved at the detector. The energy resolution depends on the finite vertical beam size at the detector and on the detector resolution itself, and is of the order of 50 MeV for energies near E_{image}. The detector consists of a camera imaging the Cherenkov light produced by positrons as they pass through a 5-cm-long air gap, delimited by two silicon wafers positioned at an angle of 45° to the beam. To measure the decelerated part of the positron spectrum, the dipole magnetic field is set to 0.4 T. By doing so, the detector field of view allows us to measure positrons with energies above 9 GeV.

For the accelerated portion of the energy spectrum presented in Fig. 2a and b, asymmetric Gaussian functions are fitted to the spectral charge density for energies above 24 GeV and 25.25 GeV, respectively. The fits are then used to calculate the number of accelerated positrons and the amount of energy gained by the peak. For the decelerated portion of the energy spectrum presented in Fig. 2c, the decelerated charge and the amount of energy transferred to the wake are calculated using all positrons that have lost more than 1 GeV. Further, the contribution of positrons with final energies below 9 GeV (that is, not detected) to the decelerated charge and the energy loss is determined using a linear extrapolation of the measured spectrum. This extrapolation accounts for about 10% of the total amount of energy loss.

To estimate the energy extraction efficiency, we divide the energy gained by the particles in the peak of Fig. 2a or b by the average amount of energy transferred to the wake by the decelerated particles (particles that have lost more than 1 GeV, including the extrapolation of the spectrum for energies below 9 GeV). The average energy loss is obtained using a piecewise reconstruction of the average decelerated spectrum, shown in Extended Data Fig. 5. In this reconstruction, each piece of the spectrum corresponds to the average of the ~35 spectra recorded for the given values of E_{image}, from 10.35 GeV to 20.35 GeV. The average energy loss obtained from this piecewise reconstruction is 2.54 J. It agrees within 5% with the average energy loss obtained by using only the data recorded at $E_{\text{image}} = 10.35$ GeV. The main error in this estimation of the energy extraction efficiency is related to the shot-to-shot fluctuation of the energy loss, which equals 0.3 J (r.m.s.) and leads to an uncertainty of ±5% and ±4% in the energy extraction efficiency of Fig. 2a and b, respectively.

Particle-in-cell simulations. Computer simulations were carried out with the three-dimensional particle-in-cell codes QuickPIC[20,21], OSIRIS[29] and Hybrid OSIRIS[30]. The outcomes of all three codes, in particular the plasma wake and the beam evolution as the positron beam propagated through the plasma, were compared against one another to ensure that all three gave the same results. The results of the quasi-static QuickPIC code are presented in Figs 1 and 3. This code uses the quasi-static approximation, which assumes the beam evolves slowly compared to the timescale of the plasma response (that is, the plasma electron transit time through the beam). The input bunch contains ~1.4×10^{10} positrons and has Gaussian profiles in both transverse and longitudinal dimensions, with $\sigma_r = 70$ μm and $\sigma_z = 30$ μm, and a r.m.s. normalized emittance of 200 mm mrad and 50 mm mrad in the two transverse dimensions, where the r.m.s. normalized emittance in x is defined as $\varepsilon_{nx} = (\langle x^2 \rangle \langle p_x^2 \rangle - \langle x p_x \rangle^2)^{0.5}/mc$ (with p_x the particle momentum along x and m the particle mass) and describes the area occupied by the beam in the phase space (x, p_x) and the r.m.s. normalized emittance in y is defined similarly. The beam has an energy of 20.35 GeV and zero initial energy spread. The plasma has a trapezoidal density profile with a uniform density of 8×10^{16} cm^{-3} over a 1.2-m-long region and 15-cm-long linear density up- and down-ramps on either side, corresponding to a FWHM length of 1.35 m. The simulation box tracks the beam–plasma interaction in the coordinates x, y and $\xi = z - ct$, that is, the box moves at light speed although it 'sees' the plasma and the positron bunch in the laboratory frame. The box has a size of 600 μm × 600 μm × 320 μm in the two transverse dimensions and the longitudinal dimension, respectively. The number of cells in the simulation box is 512 × 512 × 512 (about 134 million in total). A 790-frame movie of the simulation is available as Supplementary Video 1, and shows the evolution of the longitudinal and transverse fields of the wake as the positron beam propagates through the plasma.

Sensitivity to the positron bunch parameters. The positron bunch parameters such as the r.m.s. transverse and longitudinal spot sizes and the normalized transverse emittance can vary in the experiment. To check the sensitivity of the results to the positron bunch parameters, additional QuickPIC simulations were carried out. The following parameters were scanned over ranges that occur in the experiment: the bunch length σ_z, the longitudinal shape of the bunch (rise time versus fall time), the transverse beam size σ_r, and the normalized transverse emittance. Parameters were varied one at a time, while keeping all the other parameters constant. Other simulation parameters are as described in the Methods. The results of this simulation scan are presented in Extended Data Fig. 2, which shows the final positron spectrum for every parameter sets. Although the gross features of the wake (for example, the reversing of the E_z field within the rise time of the bunch) remain the same, the peak value of the on-axis accelerating field, the on-axis plasma electron density, and the loading of the longitudinal and transverse fields do vary noticeably over the range of parameters scanned in the simulations.

The effects of the scanned parameters on the beam–plasma interaction and the final positron spectrum can be summarized as follows.

Variation of σ_z and σ_r. In the experiment, the r.m.s. bunch sizes in the transverse and the longitudinal directions can vary by a factor of two. In the simulations, increasing σ_z or decreasing σ_r reduces the maximum energy gain and modifies the peak-to-shoulder ratio and the energy spread. An example of each is shown in Extended Data Fig. 2a.

Longitudinal pulse shape. The maximum energy of the positrons is sensitive to the details of the longitudinal profile of the bunch. For instance, if the bunch has an asymmetric profile with a sharper rise and a gentler fall such that the sum of the r.m.s rise and fall times is held constant (as well as the FWHM length of the bunch), the maximum energy gain increases substantially while the peak-to-shoulder ratio is somewhat reduced. For instance, as shown in Extended Data Fig. 2b (green curve), an asymmetric Gaussian bunch with a 20-μm rising edge and a 40-μm falling edge gives a peak energy gain increased by 2 GeV. The faster rise of the current profile of the bunch excites a wake of larger amplitude, which increases the energy gain.

Normalized transverse emittance. The final positron spectrum is sensitive to the input-normalized transverse emittance of the bunch, as shown in Extended Data Fig. 2c. Note that the variation of the normalized emittance is performed at constant σ_r, and therefore corresponds to a change in the width of the transverse momentum distribution of the beam. When the emittance is increased from 100×25 μm^2 to 200×50 μm^2, the decelerating field does not change, as can be seen from the nearly identical decelerated portion of the energy spectrum of the positrons, but the acceleration is substantially affected. Further increasing the emittance reduces the peak decelerating field. In all cases, the beam loading is reduced as the emittance is increased, which leads to a higher energy gain, a smaller peak-to-shoulder ratio and a larger peak energy spread. The spectrum is flat and has the highest energy gain for the highest emittance.

The energy gain and the energy spread of the charge contained in the narrow spectral peak, as well as the peak-to-shoulder ratio, depend strongly on the transverse and longitudinal properties of the positron bunch. In the experiment,

all these beam parameters can vary simultaneously, and one can therefore expect a wide range of outcomes, as observed in the experiment and shown in Fig. 2 and Extended Data Fig. 1.

28. Erikson, R. (ed.) *SLAC Linear Collider Design Handbook* SLAC-R-714 http://www.slac.stanford.edu/pubs/slacreports/slac-r-714.html (1984).

29. Fonseca, R. A. *et al.* OSIRIS: a three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators. *Lect. Notes Comput. Sci.* **2331,** 342 (2002).

30. Davidson, A. *et al.* Implementation of a hybrid particle code with a PIC description in *r–z* and a gridless description in ϕ into OSIRIS. *J. Comput. Phys.* **281,** 1063–1077 (2015).
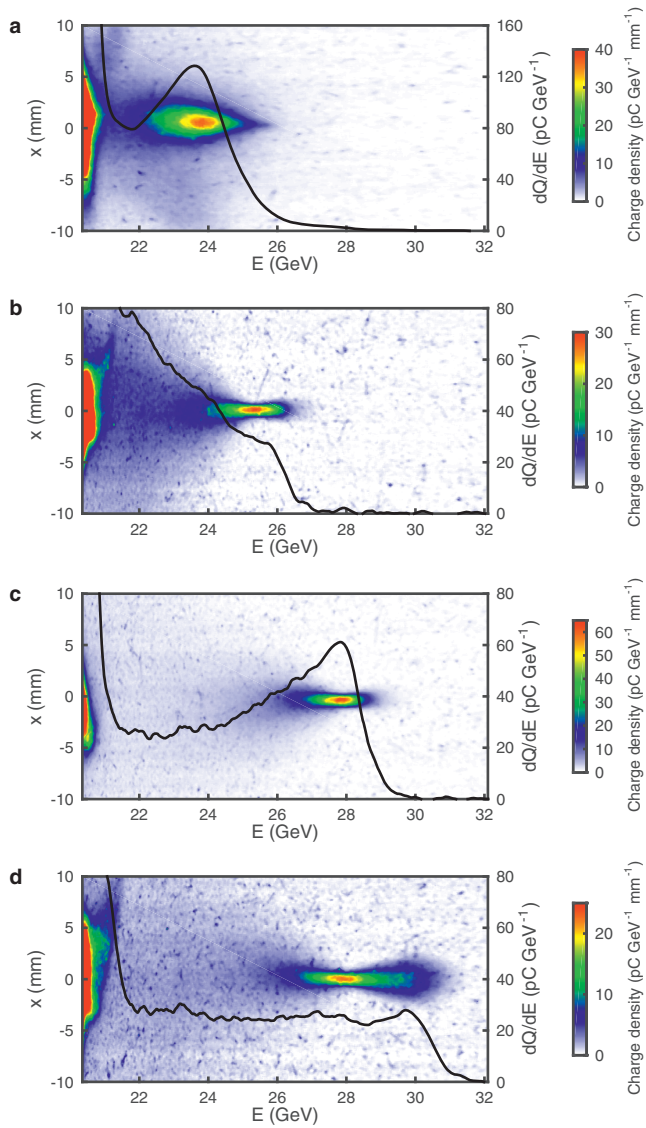
Extended Data Figure 1 | Other experimental outcomes. a–d, Various examples of the accelerated portion of the spectrum of the positrons after the interaction with the plasma, with E_{image} at 25.35 GeV (**a, b**) and 27.85 GeV (**c, d**). The charge density of the dispersed positron beam profile is shown in colour, and the spectral charge density is represented by the black solid line (right-hand scale). The spectrum is peaked with a relatively large energy spread in **a** and **c**, is continuously decreasing in **b** and is flat in **d** for the highest maximum energy gain of ∼10 GeV. Particles that do not participate in the interaction appear to be saturated at the initial beam energy of 20.35 GeV.

a

b

c
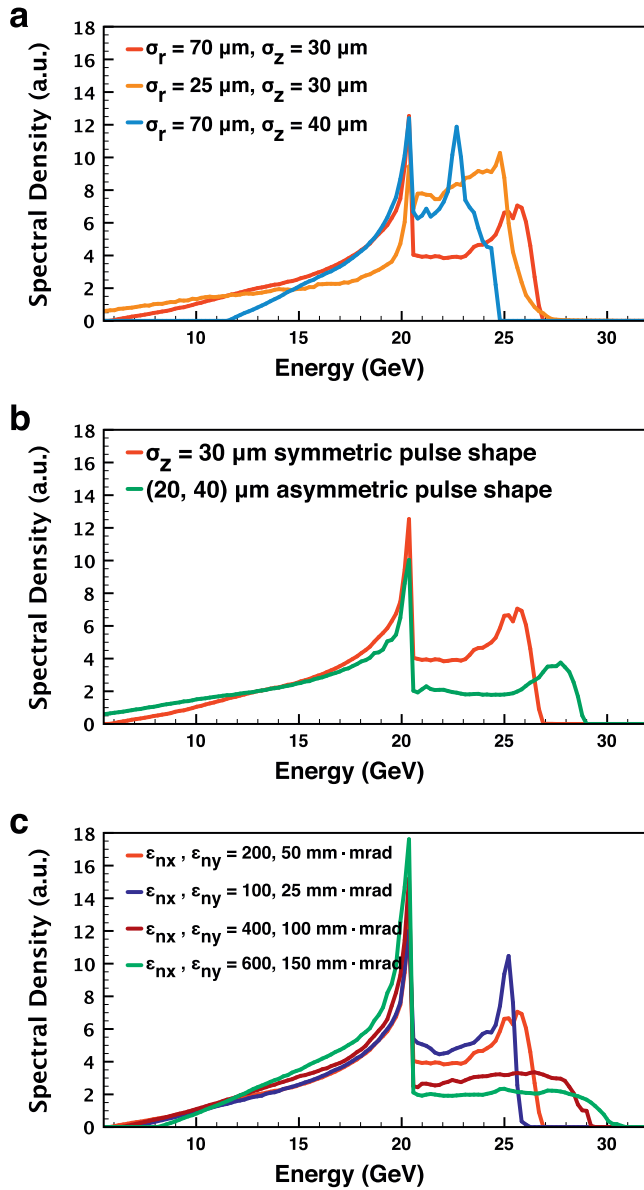
Extended Data Figure 2 │ Sensitivity to bunch parameters in computer simulations. a, Final positron spectrum from QuickPIC simulations as σ_z and σ_r are varied. **b**, Final positron spectrum as the bunch longitudinal asymmetry is varied. The red (green) curve corresponds to a longitudinally symmetric (asymmetric) bunch. **c**, Final positron spectrum as the normalized transverse emittance is varied, holding σ_r constant.
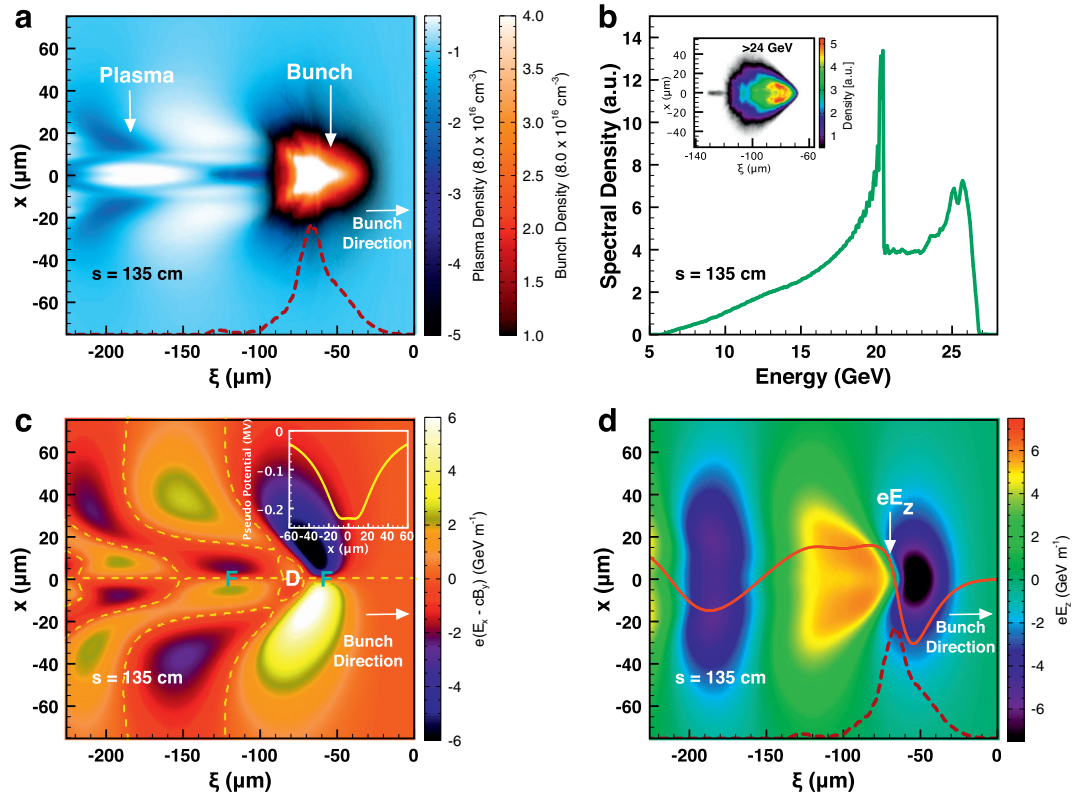
Extended Data Figure 3 | Particle-in-cell simulation results in the x–ξ plane. **a**, Electron plasma density and positron beam density in the x–ξ plane. **b**, Final spectrum of the positron bunch after its interaction with the plasma. The inset shows the projected density in the x–ξ plane of the positrons with energies greater than 24 GeV. **c**, Transverse force $F_x \approx e(E_x - cB_y)$ experienced by the positrons. The dashed lines represent the $F_x = 0$ contour. F and D indicate, respectively, the focusing and defocusing regions of the wake. The inset shows a lineout of the pseudo potential ψ (given implicitly by $\boldsymbol{F}_\perp = -\nabla_\perp e\psi$) taken at $\xi = -75\,\mu m$. **d**, Longitudinal electric field E_z. In **a** and **d**, the on-axis density profile of the positron bunch is represented in red dashed line. In comparison to the y–ξ plane shown in Fig. 3, here the transverse force is slightly defocusing close to the axis in the region $-85\,\mu m < \xi < -70\,\mu m$. The emittance being larger in x, the beam evolves towards a larger spot size along this dimension. As a result, the on-axis plasma electrons that provide the focusing force take a marginally longer time to reach the $x = 0$ plane, which leads to a small region where the force is locally slightly defocusing. Nevertheless, the positrons remain confined and guided by the pseudo potential well, as seen in the inset of **c**.
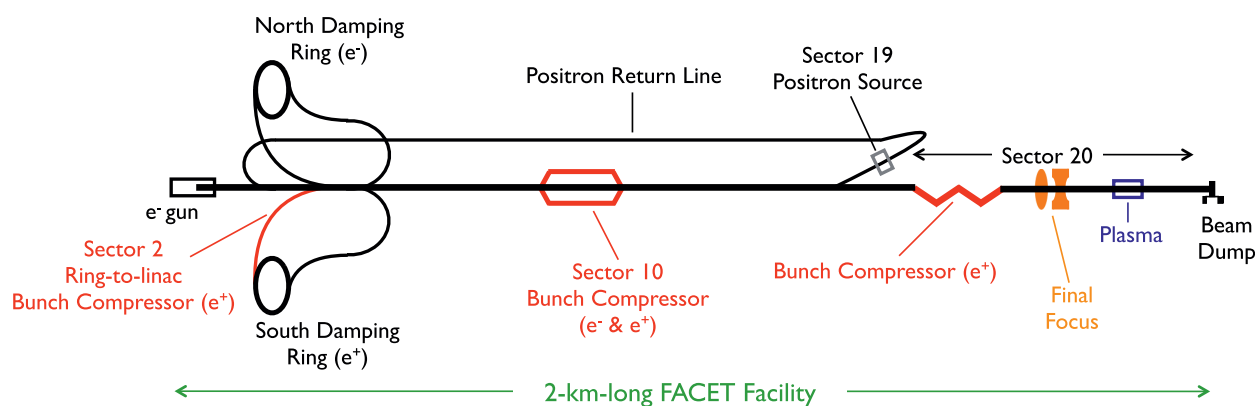
Extended Data Figure 4 | Schematic of the FACET facility. Positrons are created in the positron source of Sector 19, transported to the start of the linear accelerator by the Positron Return Line, and cooled in the South Damping ring. They are then accelerated in the linear accelerator and compressed in the ring-to-linear accelerator section, in Sector 10 and in Sector 20, and are focused by the final focus system to the entrance of the plasma in Sector 20.
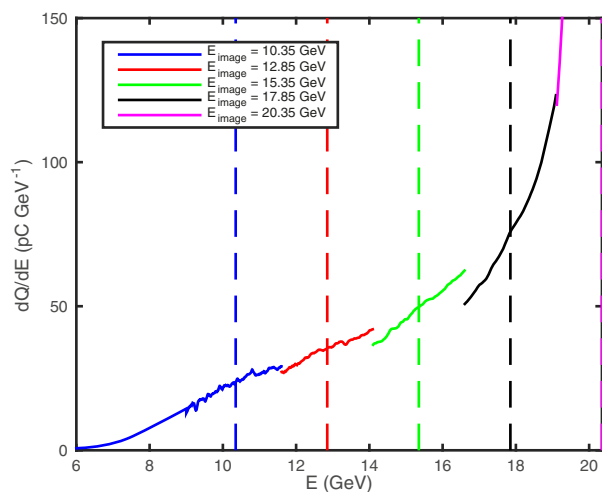
Extended Data Figure 5 | Piecewise reconstruction of the decelerated part of the average positron spectrum. Each piece of the reconstructed spectrum is the average of ~35 spectra measured at a given quadrupole energy set point E_{image}. E_{image} = 10.35 GeV is represented by the blue solid line, E_{image} = 12.85 GeV by the red solid line, E_{image} = 15.35 GeV by the green solid line, E_{image} = 17.85 GeV by the black solid line and E_{image} = 20.35 GeV by the magenta solid line. Except for the two extreme energy set-points, each piece of the reconstruction covers a 2.5-GeV energy range centred around its energy set-point. Single shot spectra are linearly extrapolated for energies below 9 GeV, and this extrapolation is accounted for in the blue curve. The energy set-points are represented by the dashed lines.

LETTER

Atomic inner-shell laser at 1.5-ångström wavelength pumped by an X-ray free-electron laser

Hitoki Yoneda[1,2], Yuichi Inubushi[2,3], Kazunori Nagamine[1], Yurina Michine[1], Haruhiko Ohashi[2,3], Hirokatsu Yumoto[3], Kazuto Yamauchi[2,4], Hidekazu Mimura[2,5], Hikaru Kitamura[6], Tetsuo Katayama[3], Tetsuya Ishikawa[2] & Makina Yabashi[2]

Since the invention of the first lasers in the visible-light region, research has aimed to produce short-wavelength lasers that generate coherent X-rays[1,2]; the shorter the wavelength, the better the imaging resolution of the laser and the shorter the pulse duration, leading to better temporal resolution in probe measurements. Recently, free-electron lasers based on self-amplified spontaneous emission[3,4] have made it possible to generate a hard-X-ray laser (that is, the photon energy is of the order of ten kiloelectronvolts) in an ångström-wavelength regime[5,6], enabling advances in fields from ultrafast X-ray spectrosopy to X-ray quantum optics. An atomic laser based on neon atoms and pumped by a soft-X-ray (that is, a photon energy of less than one kiloelectronvolt) free-electron laser has been achieved at a wavelength of 14 nanometres[7]. Here, we use a copper target and report a hard-X-ray inner-shell atomic laser operating at a wavelength of 1.5 ångströms. X-ray free-electron laser pulses with an intensity of about 10^{19} watts per square centimetre[7,8] tuned to the copper K-absorption edge produced sufficient population inversion to generate strong amplified spontaneous emission on the copper Kα lines. Furthermore, we operated the X-ray free-electron laser source in a two-colour mode[9], with one colour tuned for pumping and the other for the seed (starting) light for the laser.

We found that temporal coherence was greatly improved in the pumped copper (Cu) medium using the X-ray free-electron laser (XFEL) seeding method. For the hard-X-ray atomic laser, both narrow bandwidth and excellent longitudinal coherence are very important for ultrafast X-ray spectroscopy[10], X-ray quantum optics[11–14], and fundamental physics applications.

A major challenge in realizing atomic X-ray lasers is to create a population inversion between energy levels with large gaps between them. High-density, highly charged plasma has been used as the gain medium. In 1984, soft-X-ray lasers at wavelengths of about 20 nm were realized by neon-like selenium excited by electron collisions[1], and by hydrogen-like carbon generated in plasma recombination processes[2]. However, the shortest wavelengths achieved in these schemes were approximately 3 nm. Since then, considerable progress has been made with tabletop pumped soft-X-ray lasers[15].

An alternative approach is an optical pumping scheme, which directly generates a population inversion through photoabsorption[16,17]. Optical pumping very efficiently creates inversion without exciting unfavourable states and generates good spatial mode matching between the pumping source and the laser generated. A severe drawback has been the requirement of an intense, short-wavelength pump source. However, the advent of an intense XFEL light source (the SPring-8 Angstrom Compact Free Electron Laser, SACLA) has now made such realizations possible. In 2012, a neon soft-X-ray laser at 14.6 Å was produced by using sub-kilovolt XFEL pulses to pump the K-shell electrons of neon atoms[7]. Stimulated radiation from a solid surface was also observed in the extreme-ultraviolet region[18]. Here we

describe the realization of a hard-X-ray Cu Kα (photon energy $h\nu_s = 8.0$ keV) laser at a seed wavelength of $\lambda_s = 1.54$ Å by using intense XFEL pulses (pump wavelength $\lambda_p = 1.4$ Å) to pump the K-shell electrons of Cu atoms. Furthermore, we operated the XFEL source in a two-colour mode: one colour tuned for pumping and the other for seeding. We found that the excited Cu medium amplifies the XFEL seed pulse by two orders of magnitude with improved longitudinal coherence.

Experiments were performed at the SACLA beamline BL3[19], as schematically shown in Fig. 1. As a first experiment, we used single-colour self-amplified spontaneous-emission XFEL light tuned to the Cu K-absorption edge at $\lambda_p = 1.4$ Å (that is, a photon energy of $h\nu_p = 9.0$ keV) for optical pumping. The nominal pulse duration was approximately 7 fs with a bandwidth of about 40 eV. To increase X-ray intensity, we used a two-stage focusing system that can generate an ultrasmall spot[8] on a target. The target was a Cu foil of thickness 20 μm. The average diameter of the focal spot on the target in this experiment was measured to be 120 nm at full-width at half-maximum (FWHM) using a wire-scan method. The pulse energy at the target was about 40 μJ at maximum, which corresponds to a pump intensity of $I_p \approx 6 \times 10^{19}$ W cm^{-2}. To avoid the effects of damage induced by intense X-rays, we scanned the target so as to use a fresh surface for every shot. X-ray emission spectra in the forward direction were analysed with a dispersive spectrometer consisting of a flat silicon crystal (111 or 311 reflection) and a multi-port charge-coupled device detector[14,20,21].

In Fig. 2a, the spectral integrated energies E_s of the output Kα1 emission are plotted as a function of the pulse energy density E_p of
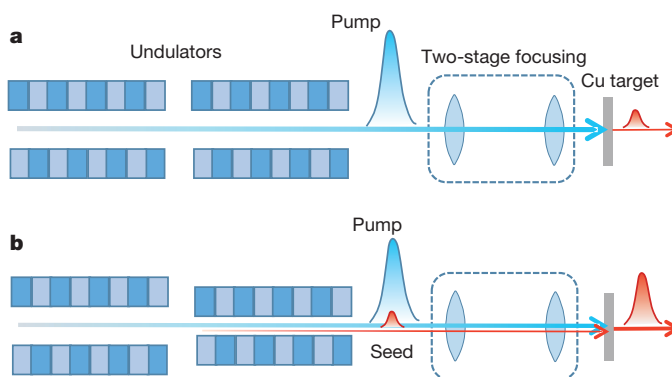

Figure 1 | Schematic drawing of experimental setup. XFEL pulses are generated with undulators in single-colour mode (the ASE experiment) (**a**) or two-colour mode (the seeding experiment) (**b**). The two-stage focusing system generates, on average, a 120-nm focusing spot. The energy ratio between the pump and the seed pulse is monitored with a transmission type spectrometer. The Kα lasing spectrum is measured with a dispersive spectrometer that combines a silicon analyser and a multi-port charge-coupled device camera.

[1]Institute for Laser Science, University of Electro-Communications, Chofu, Tokyo 182-8585, Japan. [2]RIKEN SPring-8 Center, Sayo, Hyogo 679-5148, Japan. [3]Japan Synchrotron Radiation Research Institute (JASRI), Sayo, Hyogo 679-5198, Japan. [4]Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871 Japan. [5]Department of Precision Engineering, The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan. [6]Department of Physics, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan.
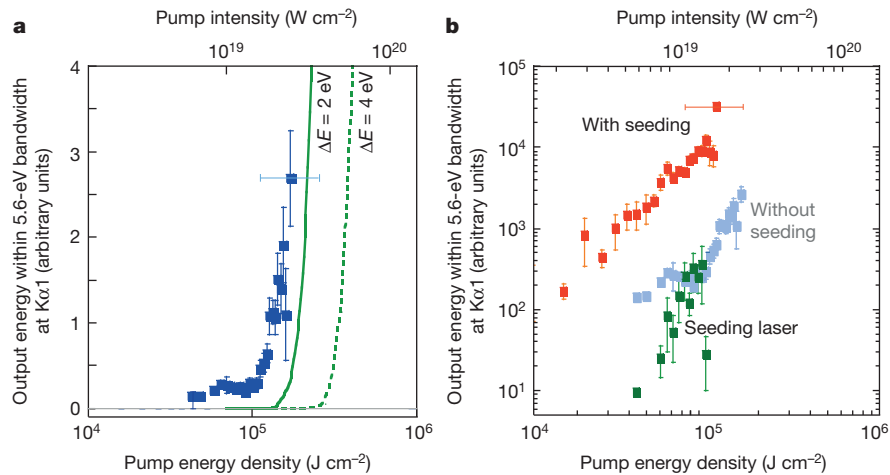
Figure 2 | Dependence of spectrally integrated output energy of Kα emission on pump pulse energy. a, ASE condition. The output energy (blue squares) of the ASE condition noticeably increases as the pump energy density increases above the threshold of 1.5×10^5 J cm^{-2} ($I_p \approx 2 \times 10^{19}$ W cm^{-2}). The green solid and dashed curves show the one-dimensional simulation results for different bandwidth conditions. **b**, Seeding condition. The output energies for the seeding case (with seeding, red squares) and those for the non-seeding case (without seeding, grey squares). Note that the latter is the same data set as plotted in **a**. Typical energies of the input seeding pulse are shown as green squares. The error bar in the vertical direction denotes the standard deviation of the measured data. The shot numbers for each condition are from several to several tens. The error bar in the horizontal direction indicates the uncertainty of the intensity of the pump pulse (see Methods for details). Both lowering of the threshold intensity of laser operation and an increase of the peak pulse energy by an order of magnitude are observed for the seeding condition.

the pumping XFEL source. The corresponding pump intensity I_p is shown on the upper axis. We observed a nonlinear enhancement of E_s increasing with the pump intensity. The enhancement becomes noticeable above the threshold of $E_p \approx 1.5 \times 10^5$ J cm^{-2} ($I_p \approx 2 \times 10^{19}$ W cm^{-2}). This dependence is typical of lasing based on an amplified spontaneous emission (ASE).

Here, XFEL pumping pulses excite Cu atoms from the ground state to the upper level (single vacancy of the K-shell electrons). This upper level is then depopulated mainly by Kα emission, Kβ emission, and Auger decay. When the pumping rate becomes higher than the total decay rate, population inversion occurs and hence lasing. First, we estimate the required intensity I_r of the pump pulse for Kα lasing, which is given by:

$$I_r = N_u D_u h\nu_p L \qquad (1)$$

where N_u is the density of the atoms in the upper level, $h\nu_p$ ($= hc/\lambda_p$) is the energy loss due to the decay, and L is the gain length[22]. D_u is the decay rate of the upper level of the laser transition, which includes the spontaneous transition probability A_{ul} and the radiative decay. Here we assume $D_u \approx 3A_{ul}$. The value of A_{ul} is a function of the stimulated cross-section σ_{stim} and the emission width $\Delta\lambda$ and $N_u\sigma_{stim}$ are replaced by the small signal gain G, the inversion factor F and the branching ratio for Kα emission from the upper level R. We can thus rewrite equation (1) as:

$$I_r = 24\pi \frac{(GL)hc^2 \left(\frac{\Delta\lambda}{\lambda}\right)R}{\lambda^3 \lambda_p F} \qquad (2)$$

(see Methods for details). Assuming $GL \approx 10$, $\Delta\lambda/\lambda \approx 5 \times 10^{-4}$, $\lambda_p \approx 0.9\lambda_s$, $R \approx 0.3$, and $F \approx 0.3$, the required intensity is given by:

$$I_r \approx \frac{7.5 \times 10^{19}}{\lambda_{K\alpha}^4} \text{W cm}^{-2} \qquad (3)$$

where $\lambda_{K\alpha}$ is the output wavelength in units of ångströms. In the present case of $\lambda_{K\alpha} = 1.54$ Å, we find the estimated threshold intensity to be $I_r \approx 3 \times 10^{19}$ W cm^{-2}, which agrees with that found in the experiment.

Next, we compared the measured dependence of output intensity on pumping intensity using a simulation. We solved the rate equations for the gain kinetics, coupled with the laser propagation equation (see Methods for details). In Fig. 3, we show a gain map in space

and time coordinates at $I_p = 3 \times 10^{19}$ W cm^{-2}. In this figure, the best gain occurs near the entrance of the Cu foil, owing to the high density of inner shell ionization. The temporal duration of the gain at each point is estimated to be approximately 1 fs, which agrees with the decay time of the upper level. Since the gain region moves as a travelling wave following the pump laser pulse, the effective gain duration of the output lasing pulse is expected to increase to about 5 fs. The temporal waveform of the Kα laser is a function of GL. At $I_p = 3 \times 10^{19}$ W cm^{-2}, the maximum GL trace is indicated by the white dotted line on Fig. 3. The total output energies are calculated by time-integration of the waveforms, which are shown as the green curves in Fig. 2a. We found
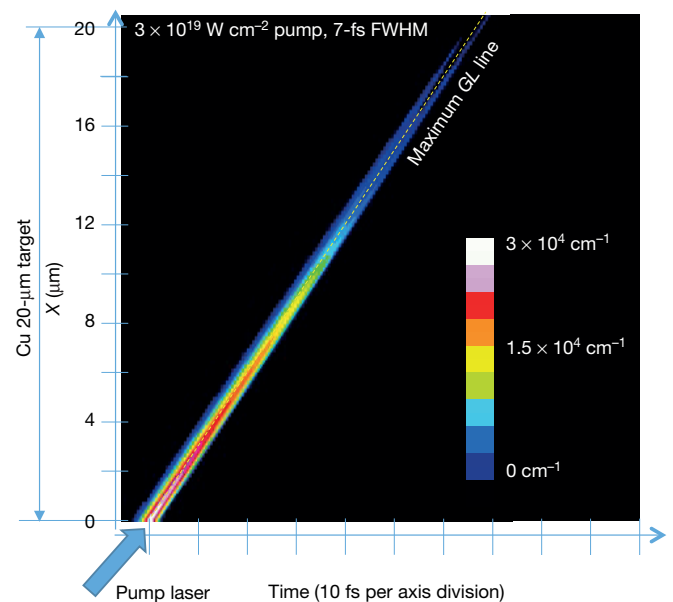

Figure 3 | Gain mapping in the time and space (X) coordinates. The gain mapping at a pump intensity of 3×10^{19} W cm^{-2} and a pulse duration of 7 fs. The maximum GL line is denoted by the white dotted line. By integrating space along with the propagation line, the output waveform of the Kα laser is calculated.

reasonable agreement (within error bars) of the threshold point and intensity dependence between the simulation and the experiment.

In Fig. 4a, we show averaged ASE spectra measured for various pump intensities. At $I_p = 1.6 \times 10^{19}$ W cm^{-2} around the measured threshold intensity only, we could observe only the Kα1 line with a typical width of approximately 2 eV, consistent with spontaneous emission of Cu Kα. Interestingly, the linewidth broadens noticeably as the pumping intensity increases. This is in contrast with what is observed in both conventional lasers and neon-gas inner-shell lasers[3], where the linewidths are decreasing or almost constant for higher intensities.

To explain this broadening, we suggest a possible mechanism based on previous studies of the spectral shape for spontaneous Kα emission of Cu (see ref. 23, for example), which also have a small tail on the low-

energy side. In these studies, the origin of the tail was explained by the presence of $3d$ vacancies, which operate as 'spectator' holes. In our case, it is likely that excitations with higher pump intensity will generate more $3d$ vacancies through several processes such as relaxation after the direct L-shell electron excitation, shifts of the energy level induced by the changed ionization states of neighbouring atoms, opacity effects[24], and excitations due to free electrons in the vicinity of the target atoms[25–27]. Increased $3d$ vacancies could then contribute to enhancing the low-energy tail and to broadening the spectral width. More detailed discussions are in the Methods.

Another noticeable feature in Fig. 4a is that, although the ratio between Kα1 and Kα2 for spontaneous emission of Cu is 2:1, reflecting the statistical weight of the levels, the ratio of the Kα1 intensity becomes much higher than this ratio in the ASE condition. This phenomenon could be caused by strong depopulation of the upper state through Kα1 emission due to the induced emission. Also, the energy states of the emitting atoms in these high-excitation conditions could be strongly affected by the neighbouring atoms (ions) in the solid[14], which may change the emission spectrum. In fact, we estimate that the ionization ratio is as high as about 10% at the threshold intensity (see Methods).

The second experiment was conducted with the XFEL source operated in the two-colour mode[9]. Here, the first section of undulators, which are tuned to the Cu K-absorption edge ($\lambda_p = 1.4$ Å), was used for optical pumping as in the first experiment, while the second section was adjusted to the Cu Kα lines ($\lambda_s = 1.54$ Å, $h\nu_s = 8.0$ keV) to produce seeding pulses. We note that the variable-gap undulator system of SACLA enables a large wavelength separation in the two-colour mode. Since the upper level has an ultrafast decay time of approximately 1 fs, we need good synchronization between the pump and seed pulses. In the two-colour operation, we could easily satisfy this requirement, because the two-colour XFEL pulses are generated from the same electron bunch. The experimental setup including the optical configuration was the same as that in the single-colour experiment shown in Fig. 1.

Figure 2b shows the spectral integrated output energies E_s of the Kα1 line, as a function of the pumping pulse energy E_p. We also plot the energy of the incident seeding pulses on the same figure. Owing to weak lasing in the XFEL system for the seeding pulse generation, the energy fluctuation of the seeding pulse is larger, but the typical seeding energy is two orders of magnitude lower than the output of the amplified Kα1 laser pulses in this experiment. We observed that the threshold of amplification is much lower than that of the ASE operation in the first experiment. In addition, we observed a decrease of the amplified gain (that is, the ratio between the energies of the amplified output pulses and the seeding input pulses) for higher pumping intensities, while the output energy does not appear to saturate (that is, the slope of the output energy versus the pump energy density decreases in Fig. 2b); this is in contrast to what is observed in conventional lasers. This characteristic can be explained by the strong absorption and small saturation intensity (see the simulation in the Methods).

Typical spectra of the output amplified pulses for various schemes are shown in Fig. 4b. For comparison, we plotted a typical spectrum of the input seed pulse, as well as that of the ASE pulse without seeding. We found large differences between these spectra. First, in the case of the seeded XFEL, both Kα1 and Kα2 can be generated, while in the ASE case Kα1 dominates, as seen in the first experiment. Second, we achieved almost single-line lasing (either Kα1 or Kα2) for the output from the seeded XFEL, while the spectrum of the input seed pulse consists of a number of narrower peaks, which change from shot to shot, within a total bandwidth of about 40 eV. Also, we observed that the width of the Kα line with seeding, 1.7 eV, is much narrower than for the spontaneous Kα line[28] or for ASE operation. This narrowing can be explained if the effective lifetime of the upper level for the amplified Kα laser line were increased. The emission width is
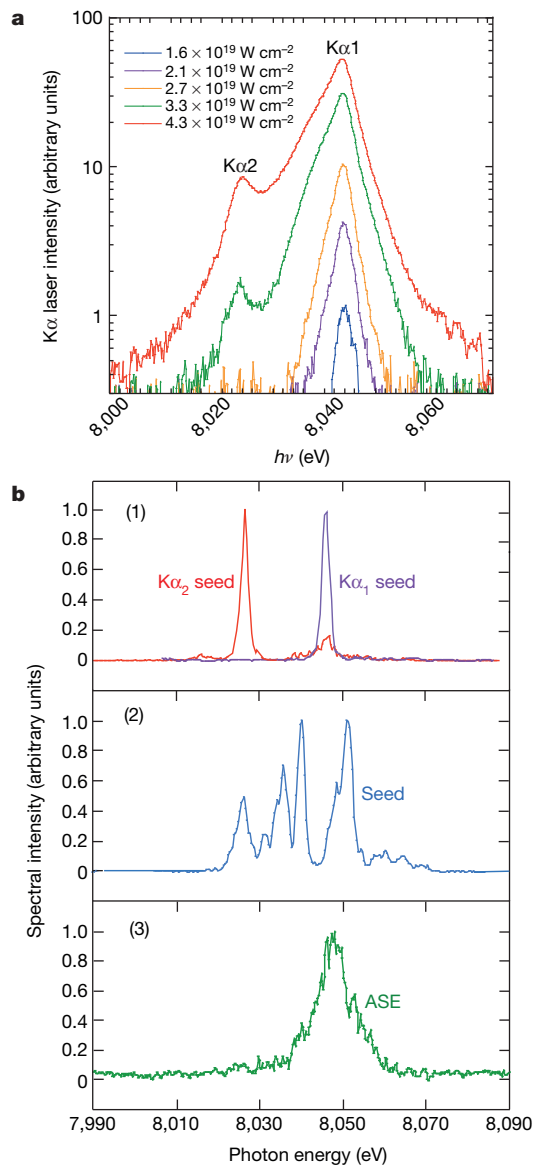

Figure 4 | Typical spectra of amplified Kα emission. a, The measured spectra of Kα ASE laser observed at different pump intensities. The broader spectral width of Kα emission is observed in higher output intensity, measured at higher pumping intensity. At $I_p = 4.3 \times 10^{19}$ W cm^{-2}, the Kα1 and Kα2 spectra overlap as a result of the broadening. **b**, Shown are (1) the seeding scheme, (2) the input seeding pulse, and (3) the ASE condition. Each spectrum is normalized by the peak spectrum intensity. Using the seeding method, single-line Kα1 or Kα2 lasers are obtained with a minimum narrowed bandwidth of 1.7 eV.

decreased by a longer interaction with co-propagation for the pump and the intense narrowed-width seeding light.

To improve the coherence of an X-ray laser, seeding with the two-colour XFEL is attractive because it gives a narrower linewidth, the ability to select the emission line (Kα1 or Kα2), and a reduced threshold intensity for laser oscillation. This new regime of X-ray laser generates a higher photon density in phase space, which will be extremely useful for the study of nonlinear and induced phenomena of X-ray interaction with matter.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Matthews, D. L. *et al.* Demonstration of a soft X-ray amplifier. *Phys. Rev. Lett.* **54,** 110–113 (1985).
2. Suckewer, S., Skinner, S., Milchberg, C. H., Keane, C. & Voorhees, D. Amplification of stimulated soft X-ray emission in a confined plasma column. *Phys. Rev. Lett.* **55,** 1753–1756 (1985).
3. Kondratenko, A. M. & Saldin, E. L. Generation of coherent radiation by a relativistic electron beam in an ondulator. *Particle Accel.* **10,** 207–216 (1980).
4. Bonifacio, R., Pellegrini, C. & Narducci, L. M. Collective instabilities and high-gain regime in a free-electron laser. *Opt. Commun.* **50,** 373–378 (1984).
5. Emma, P. *et al.* First lasing and operation of an ångstrom-wavelength free-electron laser. *Nature Photon.* **4,** 641–647 (2010).
6. Ishikawa, T. *et al.* A compact X-ray free-electron laser emitting in the sub-ångström region. *Nature Photon.* **6,** 540–544 (2012).
7. Rohringer, N. *et al.* Atomic inner-shell X-ray laser at 1.46 nanometres pumped by an X-ray free-electron laser. *Nature* **481,** 488–491 (2012).
8. Mimura, H. *et al.* Generation of 10²⁰ W/cm² hard X-ray laser pulses with two-stage reflective focusing system. *Nature Commun.* **5,** 3539 (2014).
9. Hara, T. *et al.* Two-colour hard x-ray free-electron laser with wide tunability. *Nature Commun.* **4,** 2919 (2013).
10. Young, L. *et al.* Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature* **466,** 56–62 (2012).
11. Glover, T. E. *et al.* X-ray and optical wave mixing. *Nature* **488,** 603–609 (2012).
12. Shwartz, S. *et al.* X-ray second harmonic generation. *Phys. Rev. Lett.* **112,** 163901 (2014).
13. Tamasaku, K. *et al.* X-ray two-photon absorption competing against single and sequential multiphoton processes. *Nature Photon.* **8,** 313–316 (2014).
14. Yoneda, H. *et al.* Saturable absorption of hard X-rays in iron. *Nature Commun.* **5,** 5080 (2014).
15. Suckewer, S. & Jaeglé, P. X-ray laser: past, present, and future. *Laser Phys. Lett.* **6,** 411 (2009).
16. Duguay, M. A. & Rentzepis, P. M. Some approaches to vacuum UV and X-ray lasers. *Appl. Phys. Lett.* **10,** 350–352 (1967).
17. Kapteyn, H. C., Lee, R. W. & Falcone, R. W. Observation of a short-wavelength laser pumped by Auger decay. *Phys. Rev. Lett.* **57,** 2939–2942 (1986).
18. Beye, M. *et al.* Stimulated X-ray emission for materials science. *Nature* **501,** 191–194 (2013).
19. Tono, K. *et al.* Beamline, experimental stations and photon beam diagnostics for the hard x-ray free electron laser of SACLA. *New J. Phys.* **15,** 083035 (2013).
20. Inubushi, Y. *et al.* Determination of the pulse duration of an X-ray free electron laser using highly resolved single-shot spectra. *Phys. Rev. Lett.* **109,** 144801 (2012).
21. Kameshima, T. *et al.* Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Rev. Sci. Instrum.* **85,** 033110 (2014).
22. Elton, R. C. *X-ray Lasers* (Academic, 1990).
23. Deutsch, M. *et al.* Kα and Kβ x-ray emission spectra of copper. *Phys. Rev. A* **51,** 283–296 (1995).
24. Rackstraw, D. S. *et al.* Opacity effects in a solid-density aluminium plasma created by photo-excitation with an X-ray laser. *High Energy Density Phys.* **11,** 59–69 (2014).
25. Vinko, S. M. *et al.* Creation and diagnosis of a solid-density plasma with an X-ray free-electron laser. *Nature* **482,** 59–62 (2012).
26. Ciricosta, O. *et al.* Direct measurements of the ionization potential depression in a dense plasma. *Phys. Rev. Lett.* **109,** 065002 (2012).
27. Cho, B. I. *et al.* Resonant Kα spectroscopy of solid-density aluminum plasmas. *Phys. Rev. Lett.* **109,** 245003 (2012).
28. Deutsch, M. & Hart, M. Wavelength, energy shape, and structure of the CuK at x-ray emission line. *Phys. Rev. B* **26,** 5558–5567 (1982).

Author Contributions H. Yoneda and Y.I. conceived the basic experiment. H. Yoneda, Y.I., K.N., Y.M. and M.Y. designed the research. M.Y. and T.I. are responsible for SACLA and the SACLA beamline. H.M., K.Y., H. Yumoto, H.O. and M.Y. designed and constructed the two-stage focusing system at SACLA. H. Yoneda, Y.I., K.N., Y.M. and T.K. performed the experiment. H. Yoneda and H.K. performed theoretical analyses and numerical simulations. H. Yoneda and M.Y. wrote the first draft of the manuscript with discussion and improvement from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H. Yoneda (yoneda@ils.uec.ac.jp).

METHODS

Gain, ionization ratio, and saturation intensity. The detailed dynamics of the laser operation are as follows. The small-signal gain coefficient G of a laser obeys the following formula[22]:

$$G = N_u \sigma_{stim} - N_l \sigma_{abs} \approx N_u \sigma_{stim} F \qquad (1)$$

where N_u and N_l are the number density of the upper and the lower states of the laser transition, σ_{stim} and σ_{abs} are cross-sections of simulated emission for the upper state and of absorption from the lower state to the upper state. The parameter F is the inversion factor. The cross-section σ_{stim} is given by the following formula:

$$\sigma_{stim} = \frac{c^2}{8\pi v^2} \frac{1}{\Delta v} A_{ul} = \frac{\lambda^3}{8\pi c \left(\frac{\Delta\lambda}{\lambda}\right)} A_{ul} \qquad (2)$$

where v and λ are the frequency and wavelength of the laser, c is the speed of light in vacuum, Δv and $\Delta\lambda$ are the frequency width and wavelength width of the laser, and A_{ul} is the spontaneous transition probability.

The spontaneous emission rate A_{ul} of the Kα line is given by $\sim 5 \times 10^{14}\,s^{-1}$ (ref. 24) by using the total spontaneous emission rate of Cu K emission of around $9.1 \times 10^{14}\,s^{-1}$ and the branching ratios Kα2/Kα1 ≈ 0.5 and Kβ/K$\alpha \approx 0.13$. At the measured threshold intensity of $I_p = 2 \times 10^{19}\,W\,cm^{-2}$, the peak small signal gain of the Kα1 line is estimated to be $2.2 \times 10^4\,cm^{-1}$. The length of the gain region L is similar to the absorption length of the pump X-ray laser (about 4 μm).

Next, we estimate the ratio R_i of the ionized atoms to the total atoms, which is given by:

$$R_i = \frac{\alpha I_p}{hv} \frac{\tau_{decay}}{N_{solid}} \qquad (3)$$

where τ_{decay} is the decay time of the upper level of the laser transition, α is the absorption coefficient, and N_{solid} is the initial density of the target. R_i was estimated to be 10% at the threshold intensity. This implies that the deposition profile of the pump laser penetrates beyond the cold absorption length owing to saturable absorption[14].

The saturation intensity of the laser is defined as:

$$I_{sat} = \frac{hv}{2\sigma_{stim}\tau} \qquad (4)$$

The saturation intensity of Kα emission is thus estimated to be about $6 \times 10^{16}\,W\,cm^{-2}$.

Simulation of temporal waveform using a rate equation model. As shown in Fig. 3, the gain of the Kα laser changes rapidly in time and space. To estimate the total intensity of the amplified Kα laser, we use a simulation code to solve rate equations for the populations. The code includes processes of optical pumping of XFEL, spontaneous emissions of Kα1, Kα2 and Kβ, Auger processes of KLL and KMM, induced emissions (and absorptions) of Kα1, Kα2, and Kβ, and additional excitations of L shell electrons from the upper level of Kα emission. The rate of electron collisional ionization of K and L shell electrons is lower than these processes so we neglect it. Hydrodynamic motion is a much slower process than these atomic transitions so we also neglect that. In this simulation, we assume that the linewidth of the transition is same as the natural width ($\Delta v/v = 2.4 \times 10^{-4}$) for the non-seeded case and $\Delta v/v = 1.0 \times 10^{-4}$ for the seeded case. The output energy simulated for the ASE condition with a natural bandwidth ($\Delta E = 2\,eV$) is shown as the green solid curve in Fig. 2a. The agreement between the experiment and the

simulation was reasonable, while the threshold intensity for the simulation is about two times larger than that for the experiment.

Using the rate equation model, we calculate the temporal behaviour of the output beam, the gain, and each population of the transitions. The waveforms of the seeded output pulses are shown in Extended Data Fig. 1a. In this calculation we fixed the pump waveform, which has a double-Gaussian profile with a pulse duration of 7 fs (FWHM) and a peak intensity of $2 \times 10^{19}\,W\,cm^{-2}$, which is equal to the measured threshold intensity. We also assume that the seeding pulse has a similar waveform but different peak intensity. The ratio of the peak intensities between the output and input pulses was changed from 10^{-8} to 10^{-3} in this calculation. As the input seeding energy increases, the peak intensity of the output amplified pulse is approaching saturation while the duration of the pulse constantly increases. As a result, the total pulse energy integrated along the temporal direction steadily increases, which agrees with our experimental observation.

Evaluation of the intensity on the target. We evaluated the intensities of the pump XFEL pulse on the target by using parameters of the pulse energy in the focal spot, focal spot size, and the pulse duration. The pulse energy was monitored by using calibrated photodiodes installed in different locations of the beamline. The average spot size was measured to be 120 nm (FWHM) using a knife-edge scan method. Note that the size for the single-shot exposure can be smaller than this value, because the knife-edge scan method is sensitive to small disturbances or vibrations of the optical system. The pulse duration was estimated to be 7 fs FWHM, based on a spectral spike analysis combined with a free-electron laser simulation code[20]. The uncertainties in these evaluations are shown as the horizontal error bars in Fig. 2.

Analysis of ASE spectrum. As seen in Fig. 4a, we observed notable changes of ASE spectra under high-intensity conditions. Although a full explanation of the mechanism is beyond the scope of the present study, we attempted to perform a preliminary analysis of the data by fitting the model calculation parameters to the experimentally obtained spectra.

As shown in Extended Data Fig. 1b, we decomposed the ASE spectrum measured at $I_p = 4.3 \times 10^{19}\,W\,cm^{-2}$ into several Lorentzian functions: the unshifted Kα1 and Kα2 components (labelled K1 and K2, respectively) and the shifted satellite components (labelled S1, S2, S3 and S4). We set the widths of these satellites to be equal to that for the unshifted K1 line, and the intervals between the photon energies of these satellites to be 3.2 eV. The ratio of the integrated intensities of the shifted components, normalized to that for K1, are 0.26, 0.19, 0.09, and 0.03 for S1, S2, S3, S4, respectively. The intensity ratio of the total shifted components is thus estimated to be 37%. This value becomes almost negligible for $I_p < 2.1 \times 10^{19}\,W\,cm^{-2}$.

As discussed in the main text, production of 3d vacancies is considered to be the main possible source of the changes. One of the main methods of ionizing 3d electrons is impact ionization with electrons produced by the KLL Auger process of the neighbouring atoms. In the simulations of the rate equations, 1% to 5% of atoms in the main energy extraction area ($X = 8$–16 μm inside the 20-μm target in Fig. 3) emit KLL Auger electrons at the pump intensity $I_p = 4.3 \times 10^{19}\,W\,cm^{-2}$. These electrons have kinetic energies of about 7 keV. The main ionization process for this energy is d-electron ionization with a cross-section of about $2 \times 10^{-17}\,cm^2$. A mean free path in solid Cu density is estimated to be 5.7 nm. Considering the average velocity of the electrons, the ionization rate of d-electrons is as high as $6\,fs^{-1}$ so that we expect multiple vacancy production of 3d electrons inside the X-ray-focused area. The ratio of the atoms in these highly excited states exceeds $\sim 40\%$, which is comparable to the ratio of the shifted components in the measured spectrum.
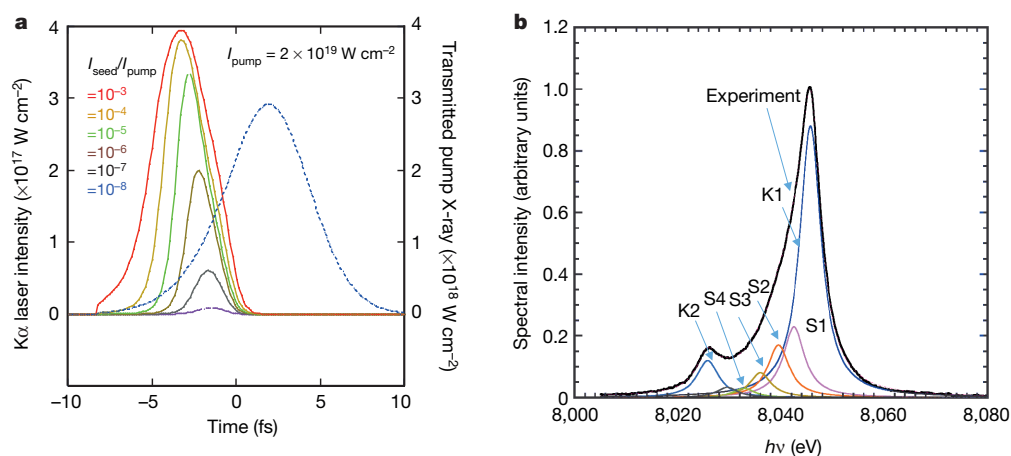
Extended Data Figure 1 | Calculated waveform and decomposition analysis of the measured Kα spectrum. a, The waveforms of the seeded output pulses at different ratios of the peak intensities between the output and input pulses. The waveform of the pump pulse is shown as a reference. **b**, A measured spectrum at $I_p = 4.3 \times 10^{19}$ W cm^{-2} and its decomposition into several Lorentzian functions. The lines labelled K1 and K2 are the unshifted components of Kα1 and Kα2, respectively. The lines S1 to S4 that have the same widths as K1 are the shifted components with a constant interval 3.2 eV in the energy shift. The total ratio of the shifted components is estimated to be about 37%.

LETTER

Substitutional doping in nanocrystal superlattices

Matteo Cargnello[1]*†, Aaron C. Johnston-Peck[2]*, Benjamin T. Diroll[1]*, Eric Wong[3], Bianca Datta[4], Divij Damodhar[4], Vicky V. T. Doan-Nguyen[4], Andrew A. Herzing[2], Cherie R. Kagan[1,4,5] & Christopher B. Murray[1,4]

Doping is a process in which atomic impurities are intentionally added to a host material to modify its properties. It has had a revolutionary impact in altering or introducing electronic[1,2], magnetic[3,4], luminescent[5,6], and catalytic[7] properties for several applications, for example in semiconductors. Here we explore and demonstrate the extension of the concept of substitutional atomic doping to nanometre-scale crystal doping, in which one nanocrystal is used to replace another to form doped self-assembled superlattices. Towards this goal, we show that gold nanocrystals act as substitutional dopants in superlattices of cadmium selenide or lead selenide nanocrystals when the size of the gold nanocrystal is very close to that of the host. The gold nano-crystals occupy random positions in the superlattice and their density is readily and widely controllable, analogous to the case of atomic doping, but here through nanocrystal self-assembly. We also show that the electronic properties of the superlattices are highly tunable and strongly affected by the presence and density of the gold nanocrystal dopants. The conductivity of lead selenide films, for example, can be manipulated over at least six orders of magnitude by the addition of gold nanocrystals and is explained by a percolation model. As this process relies on the self-assembly of uniform nanocrystals, it can be generally applied to assemble a wide variety of nanocrystal-doped structures for electronic, optical, magnetic, and catalytic materials.

Wet chemical methods can produce ensembles of monodisperse colloidal nanocrystals[8] (in which the particles are of very similar size), with sizes that can be controlled with subnanometre precision[9]. These uniform nanocrystals can self-assemble into periodic superlattices that have the same crystal structure as atomic and molecular crystals[10], with great potential for several applications[11]. Higher complexity can be obtained when two (or more) sets of monodisperse particles with different sizes or shapes are assembled into binary nanocrystal super-lattices (BNSLs) with tunable stoichiometry and crystal structure[12]. These metamaterials show emergent properties that are derived from the nanoscale interactions between the building blocks, which can be chosen from many sizes, shapes, and chemical compositions[13]. However, in BNSLs—as in binary compounds—the stoichiometry and crystal structure are interdependent and are constrained by the specific particle sizes and size ratios.

In contrast to BNSLs, in which the size difference is the driving force for the ordered assembly, two sets of monodisperse nanocrystals with very similar sizes should produce assemblies in which the two components randomly occupy the same sites of a superlattice, thus achieving substitutional doping of the superlattice. We explore this hypothesis by preparing monodisperse gold (Au) and cadmium selenide (CdSe) nanocrystals that have similar (roughly 5.5-nm) diameter (excluding the organic ligands) and size dispersions smaller than 6%, which means that each nanocrystal differs from another by no more than 6% of its size (Extended Data Fig. 1). In this method, Au and CdSe nano-crystal dispersions in hexane are mixed and allowed to self-assemble

on a diethylene glycol surface according to a well developed liquid–air interface method[14]. Ordered nanocrystal monolayers are initially tar-geted, because of their simplicity relative to thick superlattices, for an unambiguous structural characterization of the doped films.

A representative high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) image of an Au-nanocrystal-doped CdSe nanocrystal monolayer is shown in Fig. 1A, revealing that ordered, hexagonal close-packed arrays are formed. Owing to the large difference in atomic number between Au ($Z = 79$) and Cd ($Z = 48$), Au nanocrystals appear much brighter than CdSe nanocrys-tals under HAADF-STEM, while the image contrast is the opposite in bright-field transmission electron microscopy (TEM) mode (Fig. 1B). Superlattice monolayers extend to several square micrometres in area. It is evident at high magnification that local hexagonal packing is maintained and that no point defects are introduced by the Au nano-crystal dopants (Fig. 1B, C and Extended Data Fig. 2). To unequivo-cally demonstrate the presence and location of Au nanocrystals, we correlate energy-dispersive X-ray spectroscopy (EDS) compositional maps and HAADF-STEM images of the Au–CdSe nanocrystal mono-layers (Fig. 1Fa–d). Imaging of high-contrast particles in HAADF-STEM (Fig. 1Fa) reveals that lower-contrast particles are composed of Cd and Se (Fig. 1Fb) and that the bright spots are well correlated with the EDS map of Au nanocrystals (Fig. 1Fc).

The density of the dopant Au nanocrystals in the monolayer super-lattices can be tuned by varying the concentration of these nanocrystals in the starting dispersions used for self-assembly. Extended Data Fig. 3a–d shows monolayers prepared with Au nanocrystal densities of $66\,\mu m^{-2}$, $128\,\mu m^{-2}$, $183\,\mu m^{-2}$ and $317\,\mu m^{-2}$—densities that are in good accordance with the Au/CdSe ratios used in our starting solutions (see Methods for details). To assess the randomness of the process, we carried out a Voronoi analysis on the images, removing the edges, resulting in tessellations similar to that shown in Fig. 1D. The results, averaged over at least 25 different TEM images for each of the samples with the Au densities mentioned above, are shown as histograms in Fig. 1E and Extended Data Fig. 3e–g. The average particle–particle distance decreases with increasing Au density, varying from 46 nm to 41 nm to 37 nm to 25 nm. The histograms showing the distributions of Au particle–particle distances are well described by a log-normal dis-tribution, indistinguishable from a gamma distribution, as would be expected for a random tessellation[15]. The average number of edges is very close to six for all the samples. Thus the Voronoi analysis demon-strates that the Au particles behave as random substitutional dopants.

To achieve a truly random mixing of Au nanocrystals in the CdSe superlattice, the two particles must have similar size and show favour-able enthalpic contributions to mixing, obtained mainly through the chemistry of their surface ligands. The successful incorporation of Au nanocrystal dopants is realized using 1-dodecanethiol ligands. Au nanocrystals that were coated with oleylamine, but otherwise of the same size as the CdSe nanocrystals, segregated to the grain boundaries rather than being incorporated into the grains (Extended Data Fig. 4).
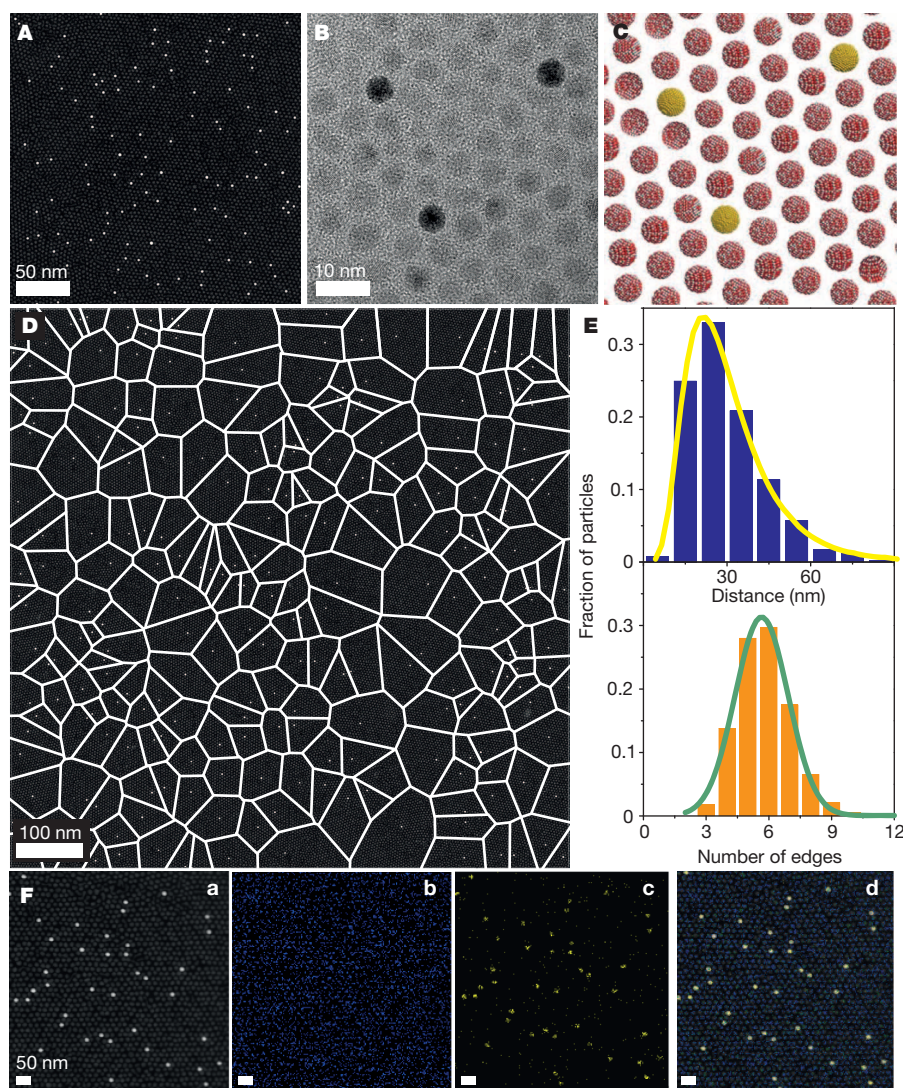
Figure 1 | **Characterization of doped monolayer superlattices.** **A**, **B**, Representative HAADF-STEM and TEM images of a CdSe nanocrystal monolayer doped with Au nanocrystals. The Au nanocrystals are seen as bright spots in **A** and dark spots in **B**. **C**, Model of the doped layer, showing the presence of Au nanocrystals (yellow) in the ordered superlattice of CdSe nanocrystals (red/grey). **D**, Voronoi diagram (white) on top of an HAADF-STEM image; Au nanocrystals are visible as small, bright dots. **E**, Histograms showing the distances between Au nanocrystals (top) and the number of edges for polygons in the Voronoi analysis (bottom); $N = 227$. **F**, **a**, HAADF-STEM mapping of an Au–CdSe nanocrystal monolayer. **b**, EDS mapping of Cd. **c**, EDS mapping of Au. **d**, Overlay EDS image.

We attribute this behaviour to the slight difference in ligand length (0.95 nm for oleylamine; 0.85 nm for dodecanethiol), the larger Hamaker constant of Au (which would dictate a more favourable interaction between Au nanocrystals[16]), and the instability of Au nanocrystals coated with oleylamine at the liquid–air interface. However, additional detailed studies are required to complete the explanation for this behaviour.

To further understand the mechanism of dopant incorporation into the lattice, we looked at the co-assembly of the same Au nanocrystals with the slightly larger (6.5-nm) lead selenide (PbSe) nanocrystals. Here, almost exclusive phase separation of the Au nanocrystals occurs, regardless of the ligand chosen for Au stabilization (Extended Data Fig. 5). Given that the deposition conditions for nanocrystal superlattice assembly and the Hamaker constants are similar for different chalcogenides (such as CdSe and PbSe), we attribute the segregation of Au nanocrystals in PbSe layers instead to size mismatch. Au nanocrystals would in fact introduce distortions into the PbSe nanocrystal superlattice, making it unfavourable. For this reason, dopants become excluded and collect together where the strain is released, for example at grain boundaries, in a process that is comparable to grain-boundary segregation of ions of different sizes in atomic-scale systems[17]. In accordance with these observations, we find that the co-assembly of slightly larger (6.8-nm) nanocrystals that consist of a gold core and a silver shell (Au/Ag) with the PbSe nanocrystals results in superlattices with long-range order, as seen for the Au–CdSe combination (Extended Data Fig. 6).

We prepared thicker films with uniform and tunable doping density by increasing the concentration of both particles in the starting solutions, resulting in ordered multilayers with Au or Au/Ag nanocrystals dispersed within the CdSe or PbSe nanocrystal superlattices (Fig. 2). High-resolution HAADF-STEM again clearly distinguishes bright Au/Ag nanocrystals that are well dispersed within three-dimensional superlattices of PbSe nanocrystals (Fig. 2a, b)[18]. We observe that Au nanocrystals are preferentially incorporated into multilayer regions (Extended Data Fig. 7), probably because of unfavourable interactions between these nanocrystals and the liquid subphase used for the assembly process. Further investigation is needed to understand this observation. Transmission small-angle X-ray scattering (TSAXS) patterns show the presence of six intense spots offset by 60°, consistent with the close-packed plane of the nanocrystal superlattice, and demonstrating that the order of a single superlattice is maintained over at least 50 μm (Fig. 2c). TEM tomography measurements on the exchanged films show that the Au/Ag dopants are dispersed throughout the superlattice volume and do not segregate to the top surface. A reconstruction of about 10^7 nm^3 is shown in Fig. 2d, together with side and top views (Fig. 2d–f and Supplementary Video), illustrating the random dispersion of the Au nanocrystals along the z axis of the films.

The presence and density of Au nanocrystal dopants drastically modifies the electronic properties of semiconductor nanocrystal superlattices. To improve the coupling between the nanocrystals[19], we displaced the long, insulating, aliphatic ligands with the more compact thiocyanate anion (SCN$^-$) through a two-step procedure that
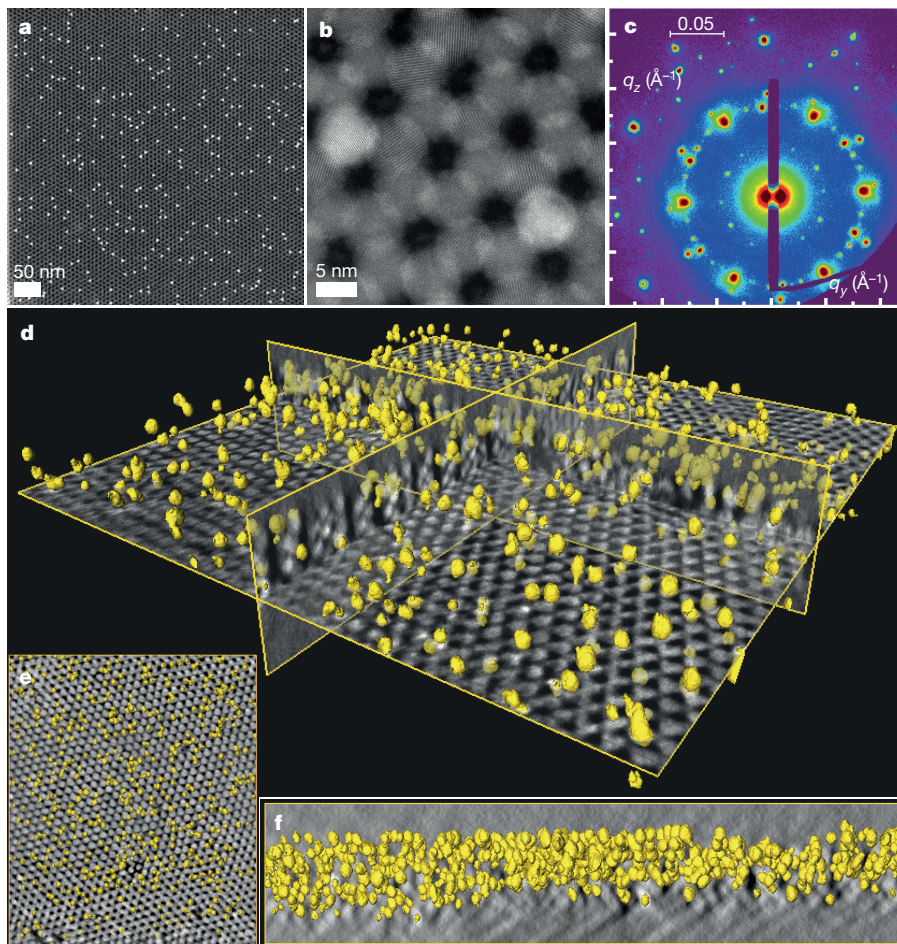
Figure 2 | Characterization of doped multi-layered superlattices. a, b, HAADF-STEM images of multilayer films of Au/Ag–PbSe nanocrystals. **c,** The corresponding TSAXS pattern shows scattered intensity along the q_z and q_y vectors. **d,** Tomographic reconstruction of a portion of the multilayer films in which Au nanocrystals (yellow spheres) are highlighted, with **e** and **f** showing the YZ and XY slice views of the same region, respectively.

preserves the superlattice structure[20]. Microscopy studies demonstrate that, despite the decrease in the interparticle distance, the superlattices remain intact (Extended Data Fig. 8). Furthermore, marked atomically

Figure 3 | Direct-current conductivity characterization of PbSe nanocrystal superlattices doped with varying amounts of Au/Ag nanocrystals. a, I–V curves, and **b,** conductivity of the nanocrystal superlattices as a function of Au/Ag doping. The blue curve in **b** shows the expectations of the Efros–Shklovskii percolation model[21]. **c,** Temperature-dependent conductivity of films of PbSe superlattices doped with 16.5% Au/Ag nanocrystals (red), and films of pure Au/Ag nanocrystals (blue). **d, e,** Schematics showing conductivity in systems below (**d**) and above (**e**) the percolation threshold in a honeycomb lattice showing the hopping path for electrons in **d** (dashed black arrow) and the direct electron transfer between Au nanocrystals in **e** (filled black arrow). The error bars in **b** show the standard deviation of 12 measurements.

coherent interfaces between nanocrystals are not observed in the electron-diffraction patterns of selected areas (Extended Data Fig. 8c)[18]. We studied the electrical conductivity of PbSe nanocrystal films doped with Au/Ag nanocrystals by laminating films on patterned substrates (Extended Data Fig. 9). Current–voltage (I–V) curves (Fig. 3a) show that the direct-current conductivity (σ) is modulated by more than six orders of magnitude, from $\sigma = 6.7 \times 10^{-6}\,\mathrm{S\,cm^{-1}}$ for pure PbSe nanocrystals, to $\sigma = 7.05\,\mathrm{S\,cm^{-1}}$ for 16.5% Au/Ag-doped PbSe superlattices (Fig. 3b). Unlike the situation that occurs with classical doping in microelectronics, modulation in the conductivity of the PbSe nanocrystal superlattice is achieved not by ionization of the Au/Ag dopant, but by percolation (described below). Other nanocrystal inclusions within superlattices may allow classical 'doping' through thermal ionization to yield n- or p-type materials.

The observed doping dependence of conductivity in Au/Ag-doped PbSe nanocrystal superlattices follows the classical behaviour observed for percolation in a random metal–dielectric composite (blue line in Fig. 3b)[21]. Percolation models predict a piecewise function to model conductivity, divided into three regions of compositional space:

$$\sigma(x) = \begin{cases} \sigma_{\mathrm{m}}(x - x_{\mathrm{c}})^{\mathrm{t}} & \text{if } x > x_{\mathrm{c}} \\ \sigma_{\mathrm{m}}\left(\dfrac{\sigma_{\mathrm{D}}}{\sigma_{\mathrm{M}}}\right)^{\mathrm{s}} & \text{if } x = x_{\mathrm{c}} \\ \sigma_{\mathrm{d}}(x_{\mathrm{c}} - x)^{-\mathrm{q}}, q > 0 & \text{if } x < x_{\mathrm{c}} \end{cases}$$

where x is the concentration of metal, x_{c} is the site percolation threshold (19.5% for hexagonal close-packed and face-centred cubic arrangements[22]), σ_{m} and σ_{d} are the conductivities of the metal and dielectric constituents respectively, and t, s, and q are exponential factors that depend on the dimensionality of the system but have no precise definition in three dimensions. The large increase of
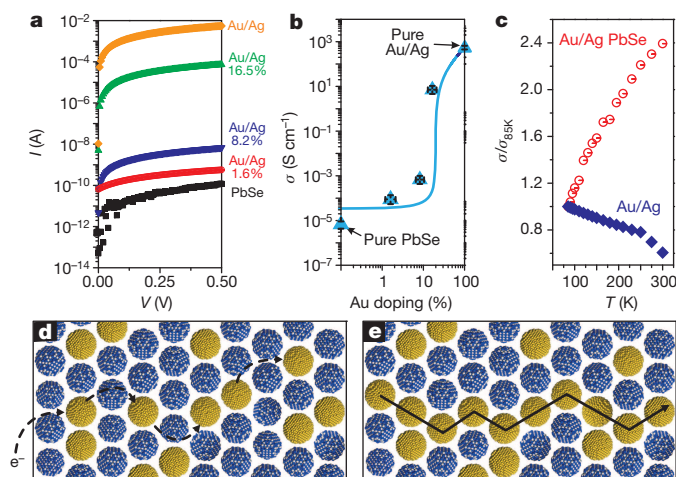
conductivity seen when doping approaches 20% is also indicative of the superlattice structure, which has a lower percolation threshold than that of glassy systems with fewer near-neighbours[23].

To investigate the microstructure of the doped superlattice further, we carried out temperature-dependent conductivity measurements of the 16.5% doping and the pure Au/Ag samples. For the 16.5% doping sample, the best single fit over the entire temperature range is a modified form of Efros–Shklovskii variable-range hopping, in which the conductivity is $\ln[\sigma(T)] \propto T^{-2/3}$ (Fig. 3c and Extended Data Fig. 10), similar to the conductivity observed for Au nanocrystal solids linked with alkane dithiols that have more than five carbons[24,25]. The observation of this behaviour in sub-percolation doped super-lattices is expected, because conductivity is constrained by hopping beyond nearest neighbours. In contrast, pure Au/Ag nanocrystal films show non-activated, metallic conductivities that increase at lower temperatures over the entire temperature range of this study (Fig. 3c), similar to the conductivities observed for Au nanocrystal films with dithiol spacers with fewer than five carbons[25], for chemically exchanged Au nanocrystals[26] or for carbon-nanotube-doped organic semiconductors[27].

In much the same way that atomic doping is essential for applications using atomic solids, we expect nanocrystal doping to show promise in many areas. The compositional variety of colloidal nanocrystals and the versatility of the self-assembly provide a way to modulate optical, magnetic, structural, and electrical properties of nanocrystal superlattices via substitutional doping. Because nanocrystals can be combined arbitrarily[28], they offer the possibility of generating composite solids with combinations of properties that are not achievable either in the bulk superlattice or in previously described binary assemblies. It is also notable that substantially diluted nanocrystal super-lattice systems can be produced without losing periodicity.

1. Erwin, S. C. et al. Doping semiconductor nanocrystals. Nature **436,** 91–94 (2005).
2. Norris, D. J., Efros, A. L. & Erwin, S. C. Doped nanocrystals. Science **319,** 1776–1779 (2008).
3. Pearton, S. J. et al. Wide band gap ferromagnetic semiconductors and oxides. J. Appl. Phys. **93,** 1–13 (2003).
4. Röder, H., Zang, J. & Bishop, A. R. Lattice effects in the colossal-magnetoresistance manganites. Phys. Rev. Lett. **76,** 1356–1359 (1996).
5. Wang, F. et al. Simultaneous phase and size control of upconversion nanocrystals through lanthanide doping. Nature **463,** 1061–1065 (2010).
6. Ye, X. et al. Morphologically controlled synthesis of colloidal upconversion nanophosphors and their shape-directed self-assembly. Proc. Natl Acad. Sci. USA **107,** 22430–22435 (2010).
7. Asahi, R. et al. Visible-light photocatalysis in nitrogen-doped titanium oxides. Science **293,** 269–271 (2001).
8. Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = S, Se, Te) semiconductor nanocrystallites. J. Am. Chem. Soc. **115,** 8706–8715 (1993).
9. Park, J. et al. One-nanometer-scale size-controlled synthesis of monodisperse magnetic iron oxide nanoparticles. Angew. Chem. Int. Ed. **44,** 2872–2877 (2005).
10. Pileni, M. P. Self organization of inorganic nanocrystals: unexpected chemical and physical properties. J. Colloid Interf. Sci. **388,** 1–8 (2012).
11. Talapin, D. V., Lee, J. S., Kovalenko, M. V. & Shevchenko, E. V. Prospects of colloidal nanocrystals for electronic and optoelectronic applications. Chem. Rev. **110,** 389–458 (2010).
12. Shevchenko, E. V. et al. Structural diversity in binary nanoparticle superlattices. Nature **439,** 55–59 (2006).
13. Urban, J. J. et al. Synergism in binary nanocrystal superlattices leads to enhanced p-type conductivity in self-assembled PbTe/Ag$_2$Te thin films. Nature Mater. **6,** 115–121 (2007).
14. Dong, A. et al. Binary nanocrystal superlattice membranes self-assembled at the liquid–air interface. Nature **466,** 474–477 (2010).
15. Tanemura, M. Statistical distributions of Poisson Voronoi cells in two and three dimensions. Forma **18,** 221–247 (2003).
16. Bodnarchuk, M. I., Kovalenko, M. V., Heiss, W. & Talapin, D. V. Energetic and entropic contributions to self-assembly of binary nanocrystal superlattices: temperature as the structure-directing factor. J. Am. Chem. Soc. **132,** 11967–11977 (2010).
17. Meledina, M. et al. Local environment of Fe dopants in nanoscale Fe: CeO$_{2-x}$ oxygen storage material. Nanoscale **7,** 3196–3204 (2015).
18. Boneschanscher, M. P. et al. Long-range orientation and atomic attachment of nanocrystals in 2D honeycomb superlattices. Science **344,** 1377–1380 (2014).
19. Fafarman, A. T. et al. Thiocyanate-capped nanocrystal colloids: vibrational reporter of surface chemistry and solution-based route to enhanced coupling in nanocrystal solids. J. Am. Chem. Soc. **133,** 15753–15761 (2011).
20. Dong, A., Jiao, Y. & Milliron, D. J. Electronically coupled nanocrystal superlattice films by in situ ligand exchange at the liquid–air interface. ACS Nano **7,** 10978–10984 (2013).
21. Efros, A. L. & Shklovskii, B. I. Critical behaviour of conductivity and dielectric constant near the metal-non-metal transition threshold. Phys. Status Solid. B **76,** 475–485 (1976).
22. Lorenz, C. D., May, R. & Ziff, R. M. Similarity of percolation thresholds on the HCP and FCC lattices. J. Stat. Phys. **98,** 961–970 (2000).
23. Rintoul, M. D. & Torquato, S. Precise determination of the critical threshold and exponents in a three-dimensional continuum percolation model. J. Phys. Math. Gen. **30,** L585 (1997).
24. Houtepen, A. J., Kockmann, D. & Vanmaekelbergh, D. Reappraisal of variable-range hopping in quantum-dot solids. Nano Lett. **8,** 3516–3520 (2008).
25. Zabet-Khosousi, A. et al. Metal to insulator transition in films of molecularly linked gold nanoparticles. Phys. Rev. Lett. **96,** 156403 (2006).
26. Fafarman, A. T. et al. Chemically tailored dielectric-to-metal transition for the design of metamaterials from nanoimprinted colloidal nanocrystals. Nano Lett. **13,** 350–357 (2013).
27. Bo, X. Z. et al. Carbon nanotubes-semiconductor networks for organic electronics: the pickup stick transistor. Appl. Phys. Lett. **86,** 1–3 (2005).
28. Overgaag, K., Liljeroth, P., Grandidier, B. & Vanmaekelbergh, D. Scanning tunneling spectroscopy of individual PbSe quantum dots and molecular aggregates stabilized in an inert nanocrystal matrix. ACS Nano **2,** 600–606 (2008).

METHODS

Synthesis of building blocks. Au nanocrystals are prepared according to ref. 29 with slight modifications. 1 mmol *tert*-butylamine-borane complex (97%, Sigma-Aldrich) is dissolved in oleylamine (80–90%, Acros Organics, 2 ml) and 1,2,3,4-tetrahydronaphthalene (tetralin, Acros Organics, 2 ml) by sonication. This solution is swiftly injected into a solution of $HAuCl_4$ (200 mg) in 20 ml oleylamine (80–90%, Acros Organics) and 20 ml 1,2,3,4-tetrahydronaphthalene (tetralin, Acros Organics) at room temperature (about 22 °C). Reduction occurs almost immediately, as shown by a change in colour from orange to dark red, but the reaction is left stirring in air for 1 h. The flask is then sealed and flushed with N_2, and the temperature is increased to 60 °C. At this point, 100 μl of 1-dodecanethiol (>98%, Sigma-Aldrich) is slowly added. At the end of the addition (after about 3 min), the flask is cooled and opened to the air. Particles are then purified by three rounds of precipitation in isopropanol/ethanol (3/1, 40 ml) and centrifugation (837.75 rad s^{-1}, 3 min), with dissolution in hexane after each step. Finally, particles are dissolved in hexane at a concentration of Au of about 5 g l^{-1}. For Ag/Au nanocrystals, 0.05 mmol of silver tetrafluoroborate were added to 2 ml Au nanocrystal solution and 1 mmol oleic acid, 1 mmol oleylamine and 1 mmol 1,2-hexadecanediol in 10 ml 1-octadecene (ODE), and the solution was heated under nitrogen gas to 180 °C for 5 min. Purification was the same as for Au nanocrystals.

Two steps are used to synthesize 5.5-nm zinc-blende CdSe nanocrystals. First, 3.5-nm CdSe cores are prepared; second, these cores are coated with additional CdSe by slow injection. The procedure is as follows. A cadmium oleate stock solution is prepared by dissolving CdO (1 mmol) in oleic acid (2.2 mmol) and 1-ODE at 270 °C until a clear solution is obtained. The concentration of Cd in the solution is 0.5×10^3 mol m^{-3}. 2 ml of this Cd stock solution are then mixed with 39 mg Se and 8 ml ODE in a 50-ml three-neck flask, and evacuated to <266.6 Pa at 110 °C for 30 min. After flushing with N_2, the flask is quickly ($\sim$50 °C min^{-1}) heated to 240 °C, and 6 ml of a solution of oleic acid/dioctylamine/ODE (1/1/2 in volume) are swiftly injected in the flask. The temperature drops upon injection and then recovers to 240 °C. 30 min after injection, the solution is heated further to 280 °C and a solution of Cd oleate stock (4 ml, 2 mmol) and Se/ODE (20 ml, 0.1×10^3 mol m^{-3}, prepared by dissolving Se in ODE at 280 °C for 4 h) is injected at 0.2 ml min^{-1} by using a syringe pump (New Era). At the end of the addition, the flask is cooled to room temperature and nanocrystals are purified by six rounds of precipitation in isopropanol/ethanol (3/1, 40 ml) and centrifugation (837.75 rad s^{-1}, 3 min) with dissolution in hexane after each step. Finally, nanocrystals are dissolved in hexane at a concentration of $\sim$25 g l^{-1}.

6.5-nm PbSe nanocrystals are prepared following ref. 30 with some modifications. PbO (1 mmol) and oleic acid (2.5 mmol) are mixed in ODE (5 ml) and evacuated (<266.6 Pa) at 110 °C until a clear solution is formed ($\sim$1 h). After flushing with N_2, the solution is heated to 180 °C and 2 ml of a solution of trioctylphosphine selenide (TOPSe) are quickly injected into the flask. The TOPSe solution is prepared by dissolving Se powder in 90% trioctylphosphine to obtain a 1×10^3 mol m^{-3} solution containing 18 μl of diphenylphosphine in a glovebox (note that diphenylphosphine is extremely flammable in contact with air; take proper precautions when handling). The temperature drops to about 150 °C, is allowed to recover to 160 °C and is left there for 20 min. After cooling to room temperature, nanocrystals are purified by performing six rounds of precipitation in isopropanol/ethanol (3/1, 40 ml) and centrifugation (837.75 rad s^{-1}, 3 min) with dissolution in hexane after each step. Finally, PbSe nanocrystals are dissolved in hexane at a concentration of $\sim$25 g l^{-1}.

Preparation of doped nanocrystal superlattices and ligand exchange. Doped nanocrystal superlattices are assembled at the air–liquid interface following an established procedure[14]. Dispersions of Au nanocrystals ($\sim$5 g l^{-1}) and CdSe or PbSe nanocrystals ($\sim$10 g l^{-1}) are mixed in a small plastic tube and the solvent evaporated using a gentle nitrogen flow. Then, 25 μl of hexane is added to redissolve the particles. This solution is then dropped onto the surface of 1.8 ml diethylene glycol (DEG) in a Teflon well ($\sim$1.5 cm × 1.5 cm × 0.9 cm). The well is covered with a glass slide to reduce the evaporation rate. After 30–60 min, the film forms and can be picked up with the appropriate substrate for characterization (for example, TEM grid, quartz piece, or patterned substrate for conductivity measurements).

A series of samples is prepared. For the initial screening of monolayers, a solution of CdSe nanocrystals diluted in hexane (1 to 2 by volume) and a solution of Au nanocrystals diluted in hexane (1 to 3 by volume). These stock solutions are mixed in the following volume ratios (CdSe/Au, μl/μl): 50/0; 50/5; 50/10; 50/15; 50/20. For the preparation of multilayers and for conductivity experiments, two solutions (CdSe or PbSe nanocrystals and Au nanocrystals) are mixed in the following volume ratios (CdSe or PbSe/Au, μl/μl): 30/0; 30/5; 30/25; 30/50; 0/50. After the film is deposited on the substrate, the film is dried under vacuum (at <3,000 Pa) overnight.

Ligand exchange is performed in two steps: *in situ* and solid exchange. The *in situ* exchange is performed by injecting 100 μl of a 0.4×10^3 mol m^{-3} solution of NH_4SCN in acetonitrile into the DEG layer while the nanocrystal film is still floating onto the DEG. The exchange is allowed to continue for 60 min or 15 min (for CdSe or PbSe, respectively) and then the films are transferred onto the substrate and evacuated (at <266.6 Pa) overnight. If the film is not thoroughly dry, it tends to delaminate during solid exchange. The solid exchange is performed by immersing the substrate with the film into a 0.4×10^3 mol m^{-3} solution of NH_4SCN in methanol for 30 s. Afterwards the film and substrate are rinsed in acetone. Fourier-transform infrared spectroscopy supports the exchange of the long oleic acid ligands with the compact SCN^-. All sample manipulations are performed in air.

Voronoi analysis. The procedure consists of extracting the signal originating from Au nanocrystals, taking advantage of the difference in contrast to filter out the CdSe nanocrystals. The Voronoi tessellations are performed using centroid information of the position of each particle, which is exported using the software ImageJ. In a typical Voronoi cell generation, the Voronoi space consists of a set of points closer to each (x, y) centroid than all other centroids. The Voronoi cell edges are generated by bisecting orthogonally the lines that connect two neighbouring centroids. This type of Delaunay triangulation, and histogram binning of the Voronoi cell areas, number of Voronoi cell edges, and centre-to-centre distances are performed using an in-house Matlab code and its "*Voronoi*" functions.

Code availability. The Voronoi tessellations generation code is freely available in a Github repository at https://github.com/vdoann/voronoi.

Characterization techniques. Substrates for electrical characterization consist of 20 nm Al_2O_3 grown by atomic-layer deposition on the surface of 250 nm thermal SiO_2 on heavily n-doped silicon wafers. Gold electrodes (10 nm) are deposited by thermal evaporation through a shadow mask. Channel lengths (L) and widths (W) range from 30 μm to 200 μm, with a constant W/L ratio of 15 for all devices.

Initial TEM characterization is performed on a JEOL JEM 1400 microscope operating at 120 kV, and on a JEOL JEM 2100 microscope operating at 200 kV. Samples are prepared using 300-mesh carbon-coated Cu grids (Ted Pella Inc.).

HAADF-STEM and EDS data are acquired using a probe-corrected Titan 80-300 microscope (FEI Co.) operated at 300 kV. The semiconvergence angle used to acquire EDS spectrum images is $\sim$21 mrad. This angle is $\sim$14 mrad or $\sim$10 mrad for acquiring HAADF images, including those in the electron tomography tilt series. Several tilt series are acquired from both the Au–CdSe and Au/Ag–PbSe samples. Generally, a tilt range from -63 ± 5 degrees to $+63 \pm 5$ degrees is sampled at increments of 1.5 degrees or 2 degrees. The tilt series are processed using Inspect3D (FEI Co.); this includes alignment and reconstruction by the simultaneous iterative reconstruction technique. Three-dimensional visualizations are performed in Avizo (FEI Co.).

Ultraviolet–visible spectra are recorded in transmission mode on a Cary 5000 spectrophotometer at 2-nm spectral bandwidth. Small-angle X-ray data are collected using a multi-angle X-ray scattering system, using a Bruker FR591 rotating anode at 40 kV and 85 mA, Osmic confocal optics, Rigaku pinhole collimation, and a Bruker HiStar Multiwire detector.

TSAXS (transmission small-angle X-ray scattering) and GISAXS (grazing incidence small-angle X-ray scattering) measurements are performed at the X9 Beamline at the National Synchrotron Light Source I (Brookhaven National Laboratory). The beam-spot size is 50-μm tall and 200-μm wide. The beam energy is 13.5 keV; the sample-to-detector distance is set to 3 m; and detection is performed using a Mar charge-coupled device detector. Typical sampling times are 10 s for GISAXS measurements and 30 s for TSAXS measurements. These conditions appear to have no negative effects on the structures measured, despite multiple measurements. Two-dimensional detector images are analysed using GISAXS Shop (https://sites.google.com/site/byeongdu/software).

Room-temperature conductivity measurements are performed using a model 4156C semiconductor parameter analyser (Agilent) in combination with contacts made from a Karl Suss PM5 probe station mounted in a nitrogen glovebox. Conductivity shows only weak dependence on the channel length, as expected for a fixed W/L ratio. Values reported are calculated from 12 independent electrode pairs from two substrates, assuming the conductive channel to have dimensions equal to the channel's cross-section.

Temperature-dependent conductivity measurements are performed under vacuum ($\sim$133.3 μPa) in a Lakeshore Cryotonics (formerly Desert Cryogenics) probe station equipped with a model 4156C semiconductor parameter analyser. For measurements in this work, samples are cooled with liquid nitrogen and measurements are taken from 85 K to 300 K. Data shown in the main text are from a single channel, but similar results are obtained from other channels.

Certain commercial equipment and materials are identified in this paper in order to specify the experimental procedure adequately. In no case does such identification imply recommendations by the National Institute of Standards and Technology nor does it imply that the material or equipment identified is necessarily the best available for this purpose.

29. Peng, S. *et al.* A facile synthesis of monodisperse Au nanoparticles and their catalysis of CO oxidation. *Nano Research* **1,** 229–234 (2008).

30. Yu, W. W., Falkner, J. C., Shih, B. S. & Colvin, V. L. Preparation and characterization of monodisperse PbSe semiconductor nanocrystals in a noncoordinating solvent. *Chem. Mater.* **16,** 3318–3322 (2004).
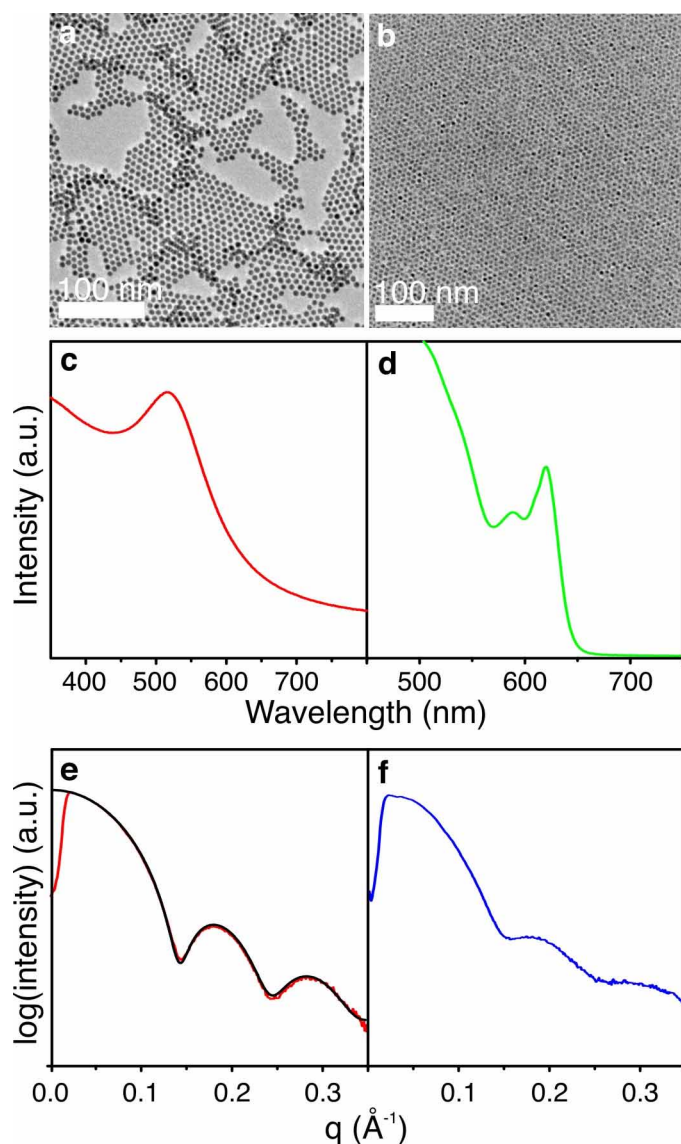
Extended Data Figure 1 | Characterization of starting building blocks.
a, **b**, Representative TEM images; **c**, **d**, ultraviolet–visible spectra; **e**, **f**, small-angle X-ray scattering (SAXS) patterns, of Au nanocrystals (**a**, **c**, **e**) and CdSe nanocrystals (**b**, **d**, **f**). For Au, the SAXS data include the fit to particle size and size dispersion (solid black trace), providing an estimated particle diameter of 6.85 ± 0.45 nm (95% confidence, including organic ligand shell).
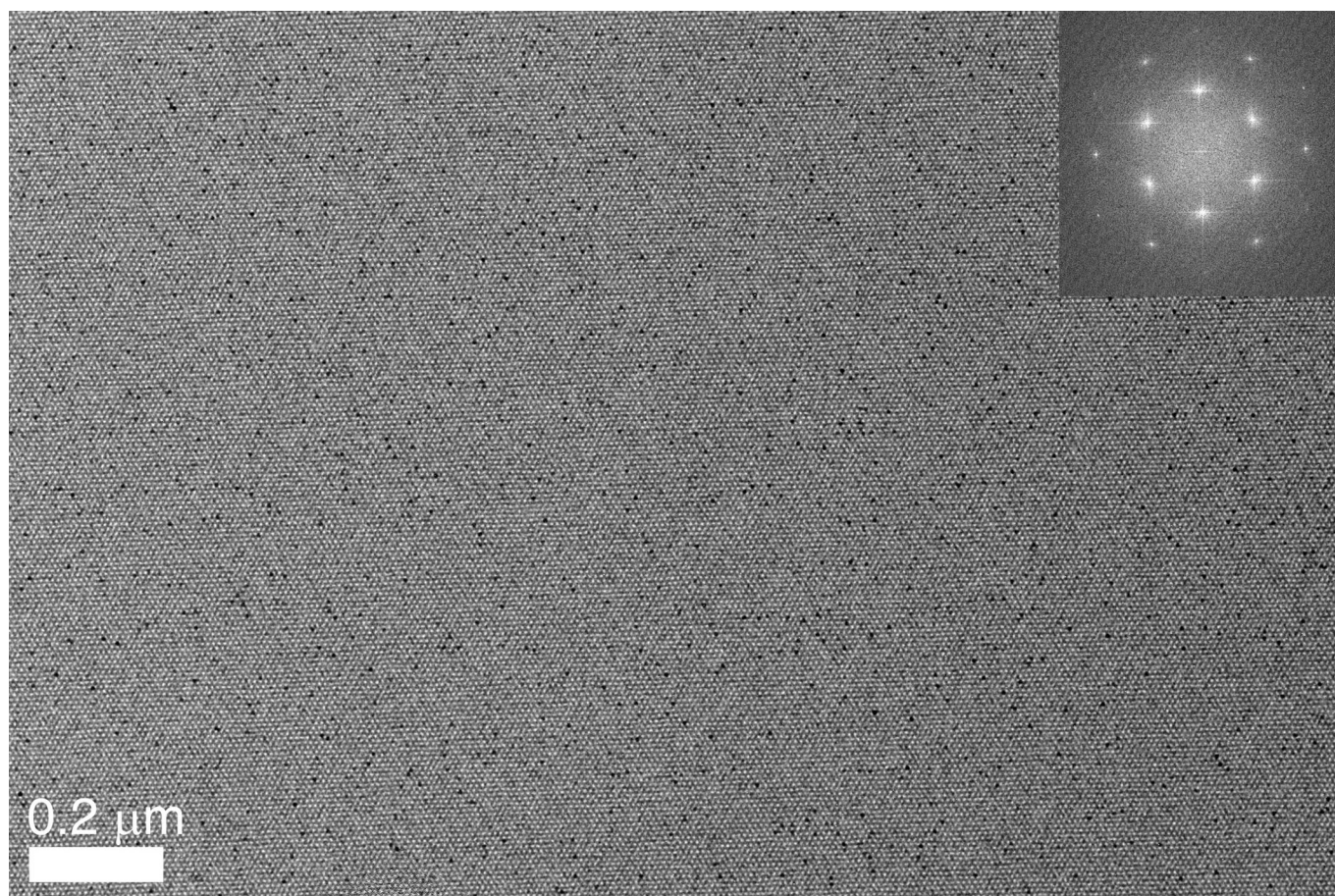
Extended Data Figure 2 | Long-range large-area ordering in doped nanocrystal superlattices. Low-magnification TEM image of CdSe superlattices doped with Au, showing the large areas (at least 5 μm²) of ordered assemblies formed onto TEM grids. Inset, digital diffraction pattern showing the hexagonal long-range ordering in the superlattices.
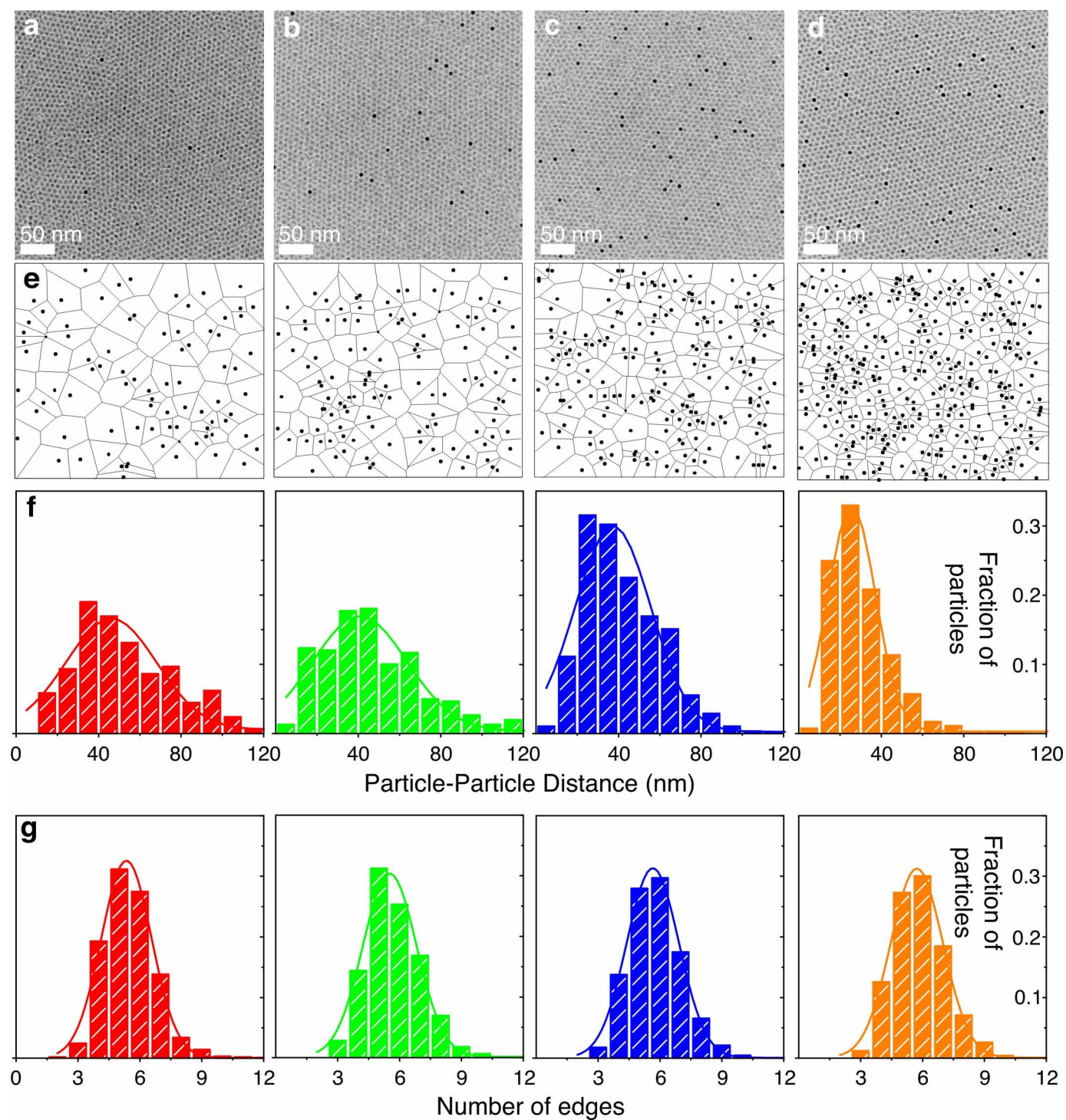
Extended Data Figure 3 | Voronoi analysis of doped monolayer super-lattices. a–d, Representative TEM images of mixed Au/CdSe ordered monolayers with different densities of Au nanocrystals: **a**, 66 nanocrystals μm^{-2};

b, 128 nanocrystals μm^{-2}; **c**, 183 nanocrystals μm^{-2}; **d**, 317 nanocrystals μm^{-2}. **e**, Corresponding larger-area Voronoi diagrams. **f**, Associated particle–particle distance histograms. **g**, Average number of edges.
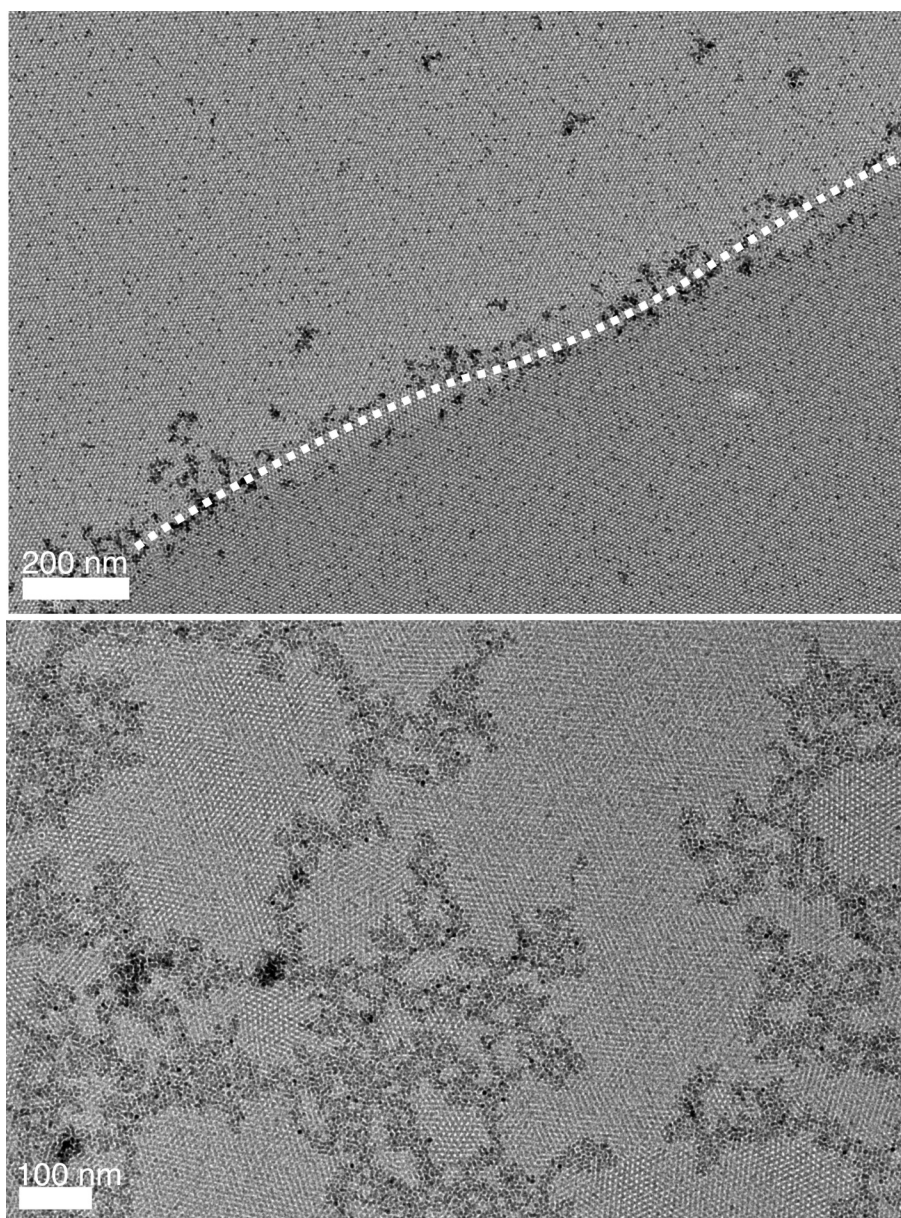
Extended Data Figure 4 | Segregation of dopants. Superlattices assembled from oleylamine-capped Au nanocrystals and oleic-acid-capped CdSe nanocrystals. Au nanocrystals tend to segregate and agglomerate at grain boundaries instead of mixing with CdSe nanocrystals. The grain boundary between two assembled regions of CdSe nanocrystals is highlighted by a dashed white line.
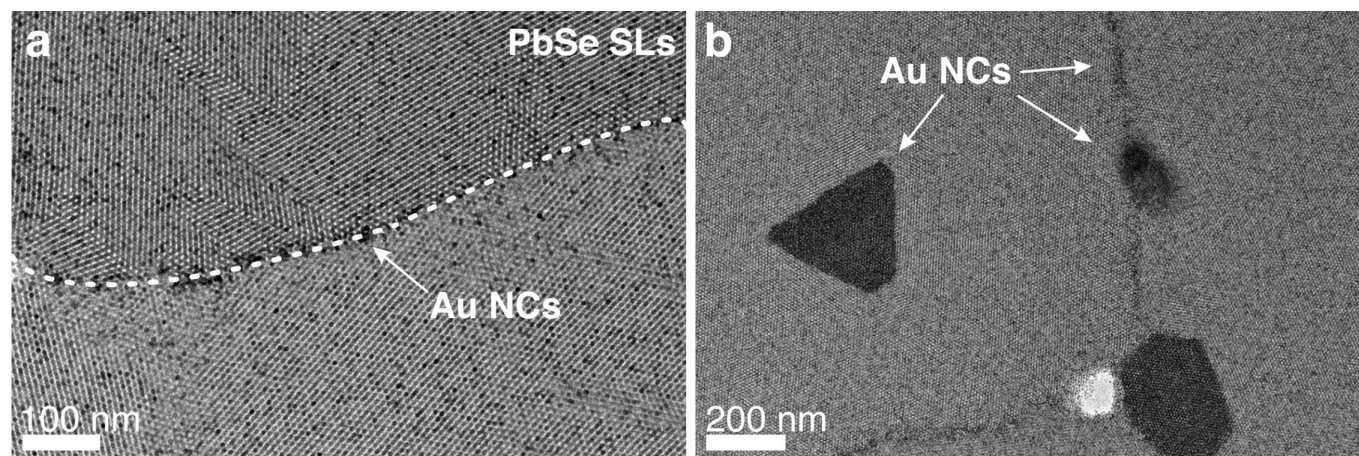
Extended Data Figure 5 | Segregation of Au nanocrystals owing to size mismatch. Assembly of 6.5-nm PbSe nanocrystals and 5.5-nm Au nanocrystals (NCs) at the liquid–air interface. Arrows point to grain boundaries or regions on top of the PbSe superlattices (SLs) where Au nanocrystals segregate preferentially.
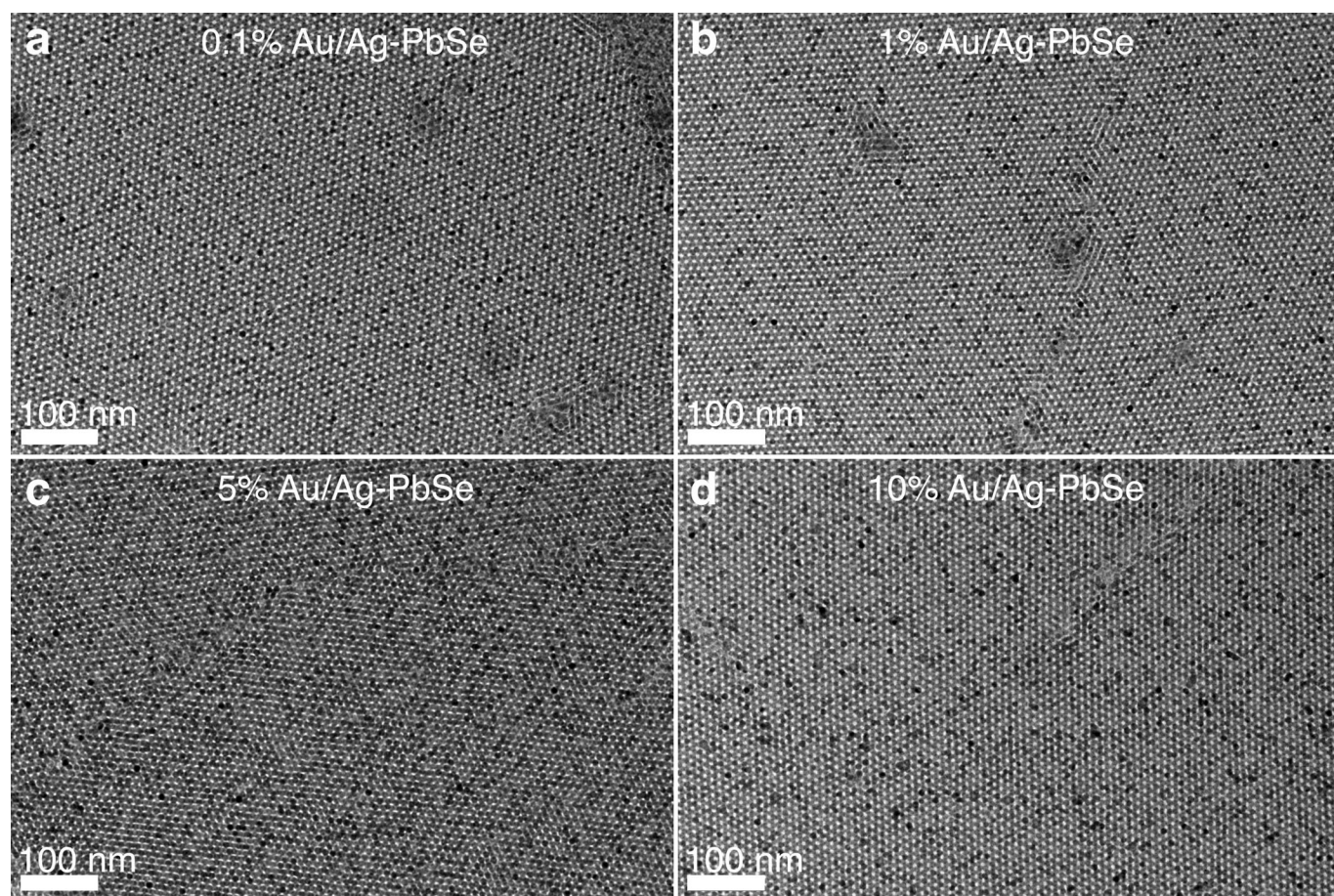
Extended Data Figure 6 | PbSe superlattices doped with Au/Ag nanocrystals. The images show the assembly of PbSe superlattices doped with different Au/Ag concentrations by volume at the liquid–air interface.
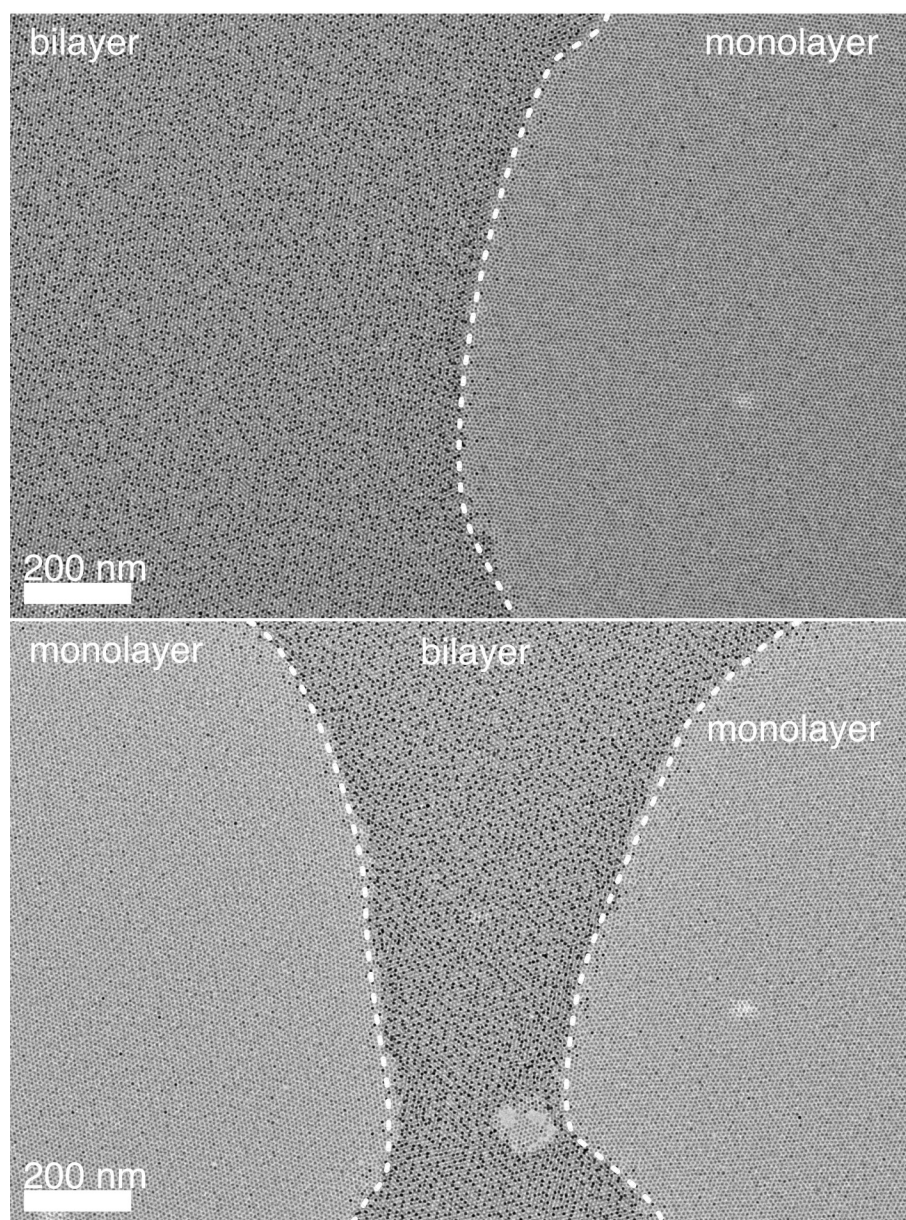
Extended Data Figure 7 | Segregation of dopants in multilayers. TEM images showing preferential segregation of Au nanocrystals to regions of CdSe multilayers rather than monolayers. Grain boundaries are highlighted with dashed white lines.
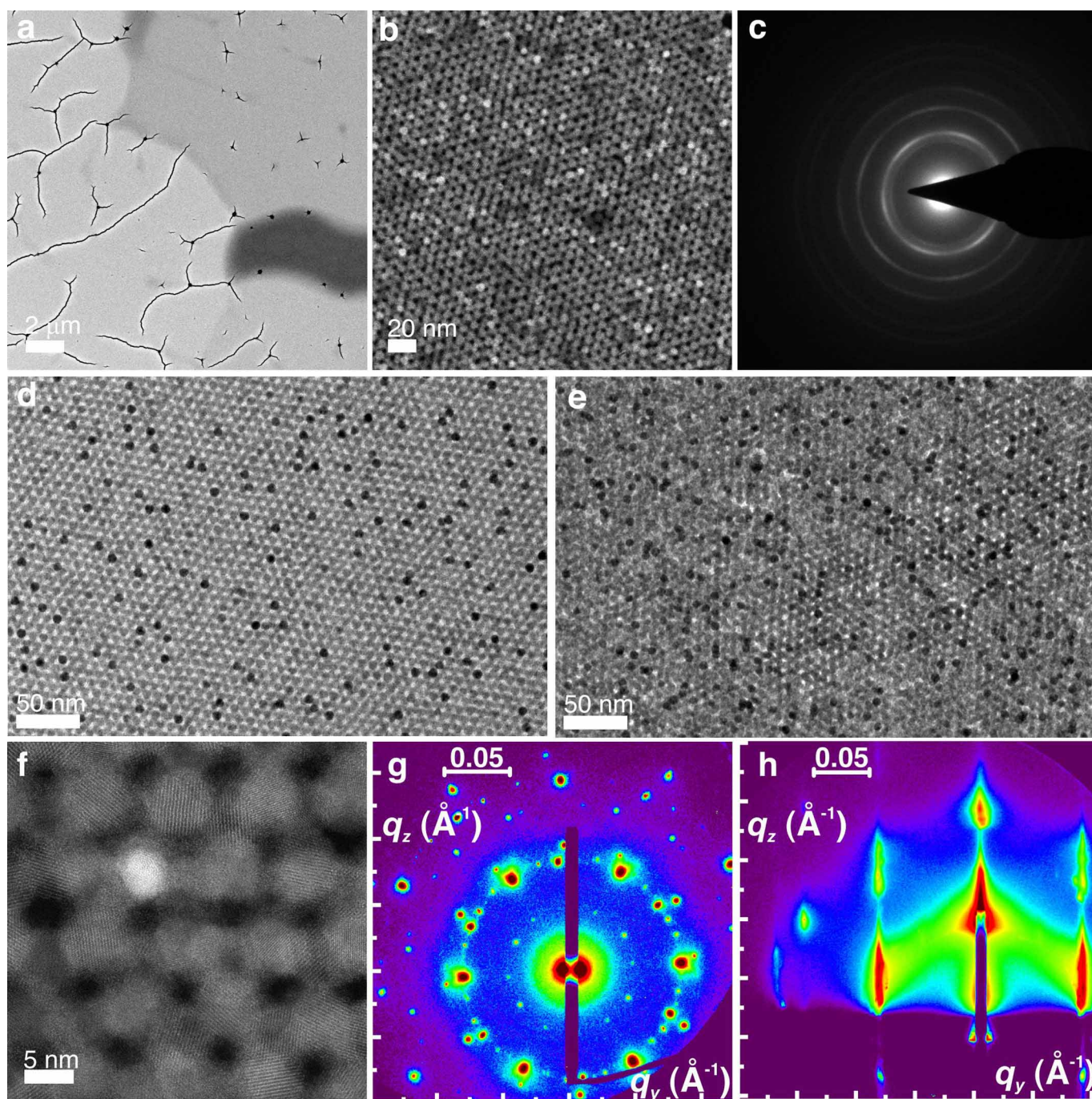
Extended Data Figure 8 | Characterization of ligand-exchanged doped superlattices. **a**, **b**, Representative TEM images of Au/Ag-doped PbSe films following ligand exchange of the long, alkyl ligands with compact thiocyanate ions. Extensive cracks are formed, especially where the grain boundaries were present. **c**, Selected area electron diffraction pattern of Au/Ag–PbSe assemblies, showing a low degree of preferential orientation of the crystallographic axes. **d**, **e**, Au–CdSe films before (**d**) and after (**e**) ligand exchange, showing that the order is preserved to a large extent, given the presence of hexagonal close-packed patches in the sample. **f**, High-resolution STEM of an Au/Ag–PbSe film after ligand exchange. **g**, **h**, TSAXS and GISAXS analysis of the films before (**g**) and after (**h**) ligand exchange.
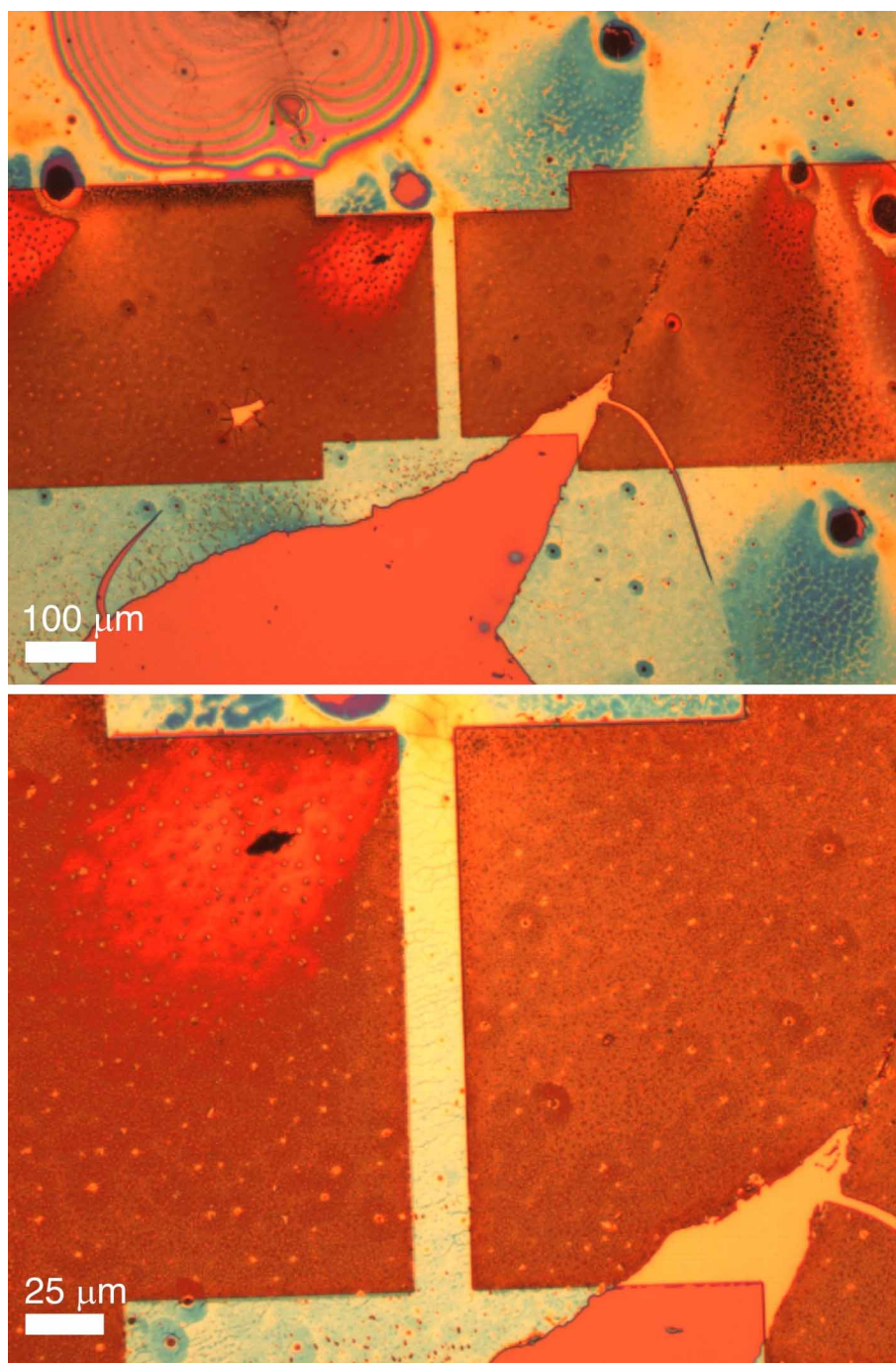
Extended Data Figure 9 | Substrates for conductivity measurements. Optical microscopy images of the patterned substrates covered with the doped superlattices for conductivity measurements.
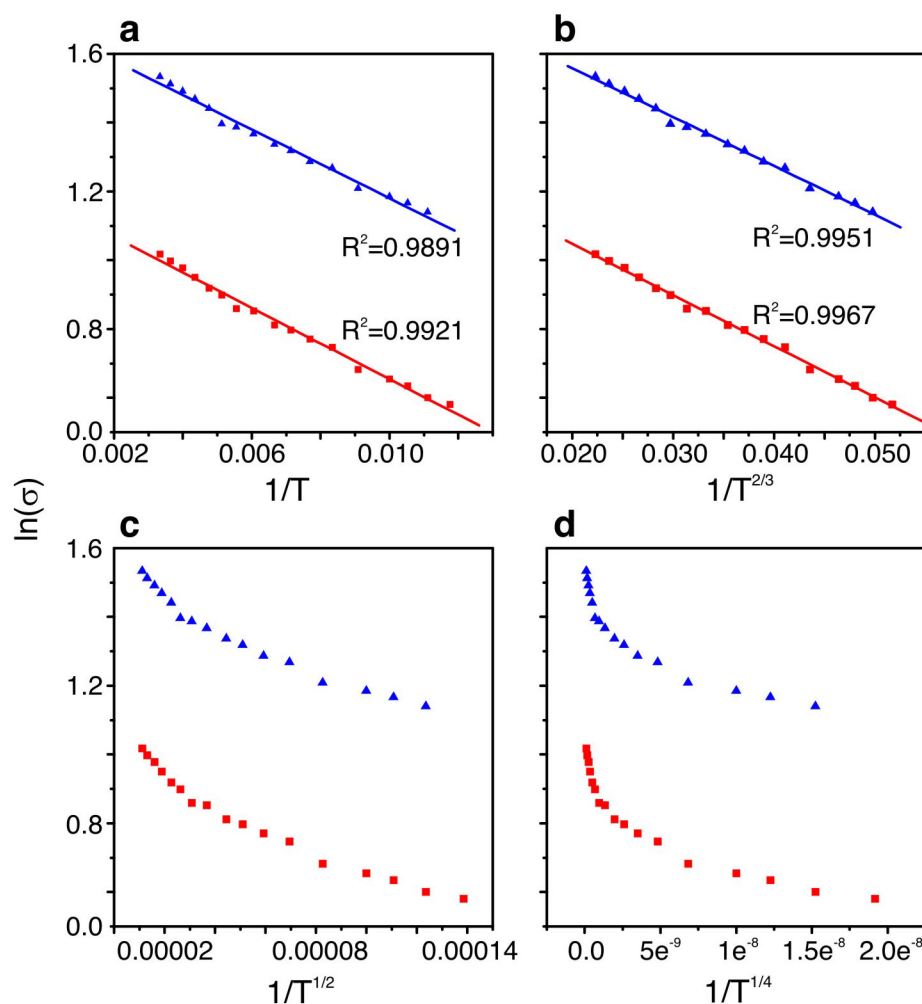
Extended Data Figure 10 | Temperature-dependent conductivity models. a–d, Fitting of different temperature-dependent conductivity models to data obtained for samples of PbSe doped with 16.5% Au/Ag. The best single fit out of several possibilities over the entire temperature range is a modified form of the Efros–Shklovskii variable-range hopping with $T^{-2/3}$ dependence, shown in **b**. The observation of this behaviour in sub-percolation doped superlattices is to be expected, because conductivity is constrained by hopping to non-nearest neighbours. The good fit obtained using the general T^{-1} expression for near-neighbour hopping, shown in **a**, not surprisingly still describes the system quite well, given that the composition is close to the percolation threshold. However, a reduced activation energy analysis also suggests that the transport behaviour is in best agreement with modified Efros–Shklovskii hopping. The original Efros–Shklovskii variable-range hopping dependence with $T^{-1/2}$ shown in **c** was found in the past to describe CdSe solids well; however, the presence of Au nanocrystal dopants clearly changes the behaviour in our system compared with the behaviour of pure quantum dot solids. The Mott variable-range hopping with a dependency of $T^{-1/4}$, shown in **d**, clearly does not describe the system, probably because of the increased density of states in the solid owing to the presence of the Au nanocrystal dopants.

LETTER

Multimetallic catalysed cross–coupling of aryl bromides with aryl triflates

Laura K. G. Ackerman[1], Matthew M. Lovell[1] & Daniel J. Weix[1]

The advent of transition-metal catalysed strategies for forming new carbon-carbon bonds has revolutionized the field of organic chemistry, enabling the efficient synthesis of ligands, materials, and biologically active molecules[1–3]. In cases where a single metal fails to promote a selective or efficient transformation, the synergistic cooperation[4] of two distinct catalysts—multimetallic catalysis—can be used instead. Many important reactions rely on multimetallic catalysis[5–10], such as the Wacker oxidation of olefins[6–8] and the Sonogashira coupling of alkynes with aryl halides[9,10], but this approach has largely been limited to the use of metals with distinct reactivities, with only one metal catalyst undergoing oxidative addition[11,12]. Here, we demonstrate that cooperativity between two group 10 metal catalysts—(bipyridine)-nickel and (1,3-bis(diphenylphosphino)propane)palladium—enables a general cross-Ullmann reaction (the cross-coupling of two different aryl electrophiles)[13–15]. Our method couples aryl bromides with aryl triflates directly, eliminating the use of arylmetal reagents and avoiding the challenge of differentiating between multiple carbon–hydrogen bonds that is required for direct arylation methods[16,17]. Selectivity can be achieved without an excess of either substrate and originates from the orthogonal reactivity of the two catalysts and the relative stability of the two arylmetal intermediates. While (1,3-bis(diphenylphosphino)propane)palladium reacts preferentially with aryl triflates to afford a persistent intermediate, (bipyridine)nickel reacts preferentially with aryl bromides to form a transient, reactive intermediate. Although each catalyst forms less than 5 per cent cross-coupled product in isolation, together they are able to achieve a yield of up to 94 per cent. Our results reveal a new method for the synthesis of biaryls, heteroaryls, and dienes, as well as a general mechanism for the selective transfer of ligands between two metal catalysts. We anticipate that this reaction will simplify the synthesis of pharmaceuticals, many of which are currently made with pre-formed organometallic reagents[1–3], and lead to the discovery of new multimetallic reactions.

We envisaged that a general solution to the cross-Ullmann reaction could be realized through multimetallic catalysis if, first, each of the two catalysts activated only one of the two substrates (1 and 2 in Fig. 1); second, selective transmetallation (the transfer of ligands from one metal to another) could be achieved (overlap of circles in Fig. 1); and third, the catalysts were redox compatible[18]. Based on Hayashi's work[19], we chose an electron-rich palladium(0) (Pd) catalyst with a bidentate phosphine ligand, 1,3-bis(diphenylphosphino)propane (dppp), to react preferentially with aryltrifluoromethylsulfonate esters (aryl triflates) over aryl bromides[19]. From our own studies[20], we chose a bipyridine (bpy) nickel(0) (Ni) catalyst to react selectively with aryl bromides in preference to aryl triflates.

We began our investigation of this multimetallic system by combining a 1:1 mixture of bromobenzene (substrate 2) and 4-methoxy-phenyltriflate (substrate 1) in the presence of both of the catalysts and a zinc reducing agent (Fig. 2b). Remarkably, we observed a high selectivity for formation of the cross-product, 4-methoxybiphenyl (product

3; 70% yield), in preference to the dimers biphenyl (product 4; 17%) and bianisole (product 5; 10%). When each catalyst was allowed to react independently with the two aryl electrophiles, however, no such 'cross-selectivity' was observed. (Bpy)NiBr$_2$ formed exclusively biphenyl (4 in Fig. 2a) from bromobenzene (2) before reacting with 4-methoxyphenyltriflate (1) (Fig. 2c), while (dppp)PdCl$_2$ was unreactive under these conditions, consuming only a minimal amount of substrate and forming trace amounts of benzene, anisole, bianisole (5), and product (3) (Fig. 2d). Only when the two catalysts were combined did a selective reaction occur.

To ensure that the cross-product observed was the result of synergy between the two proposed metal catalysts, we examined the reactivity of alternative catalysts that could be formed in situ via exchange of the supporting ligands dppp and bpy[21]. Our results (Table 1, entries 5, 7–9) demonstrate that the two alternative catalysts, (dppp)NiBr$_2$ and (bpy)PdCl$_2$, are poorly reactive and form primarily biphenyl (4) rather than cross-product (3). These findings are consistent with product arising from our proposed mechanism, and with symmetrical biaryl products arising from 'mismatched' metal–ligand combinations. Interestingly, when all of the reagents were added together at the beginning of the reaction, the results were nearly identical to reactions in which pre-formed catalysts were used (Table 1, entry 3 versus entry 10).

Examination of different ligands for both nickel and palladium revealed that while many nitrogen-based ligands were effective, relatively few bisphosphines worked well. Nickel complexes derived from bidentate ligands (2,2′-bipyridine, 1,10-phenanthroline, 2-pyridylimidazole)
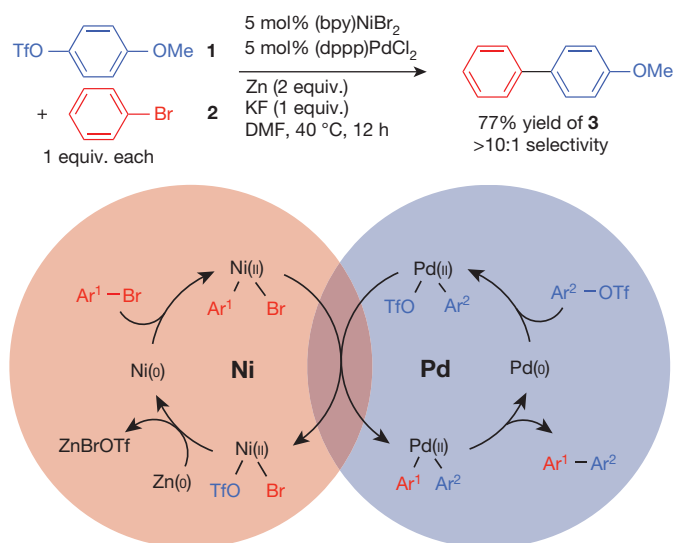
Figure 1 | A general cross-Ullmann reaction catalysed by a combination of nickel and palladium. 1, 4-methoxyphenyltriflate; 2, bromobenzene; 3, the main cross-product, 4-methoxybiphenyl. Although a nickel(0/II) and palladium (0/II) cycle is depicted, alternative mechanisms are also possible, such as a nickel(I/III) cycle[30]. OTf, triflate; DMF, N,N-dimethylformamide (solvent).

[1]Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, USA.
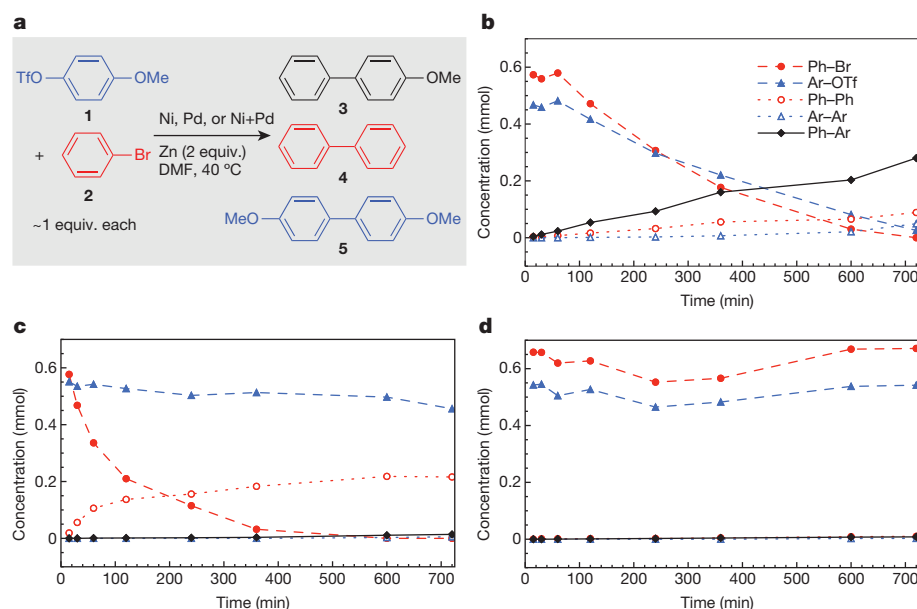
Figure 2 | Selectivities of nickel and palladium catalysts. a, The substrates and three possible products. **1**, 4-methoxyphenyltriflate (Ar-OTf); **2**, bromobenzene (Ph-Br); **3**, 4-methoxybiphenyl (Ph-Ar); **4**, biphenyl (Ph-Ph); **5**, bianisole (Ar-Ar). **b**, (dppp)PdCl₂ and (bpy)NiBr₂ (10 mol% each); **c**, (bpy)NiBr₂ (10 mol%); and **d**, (dppp)PdCl₂ (10 mol%) were used along with approximately 0.5 mmol of each starting material. Concentrations of products were determined by gas chromatography analysis, corrected with an internal standard. The low mass balance in **c** is due to the formation of benzene (0.13 mmol) from hydrodehalogenation of substrate **2**.

or a tridentate ligand (2,2′:6′,2″-terpyridine) all formed reasonably selective multimetallic catalytic systems with (dppp)PdCl₂ (47–69% yield; Extended Data Fig. 1). By contrast, only two bisphosphine ligands besides dppp were effective: 1,2-bis(diphenylphosphino)benzene (56% yield) and (Z)-1,2-bis(diphenylphosphino)ethylene (53% yield). The performance of dppp might arise from improved desired reactivity with palladium or from diminished undesired reactivity with nickel. We also expected the ratio of nickel to palladium to be important to selectivity, considering the vastly different reactivities of the individual complexes, but we found that reactions using 2/1, 1/1, or 1/2 Pd/Ni ratios provided similar yields and selectivities (Table 1,

entries 11 and 12 versus entries 2 and 3; see also Supplementary Information, Chart S6). Together, these results suggest that the transmetallation step between the two catalysts is inherently selective.

In many cases, the ratio of substrates also had minimal effect on the final yield of product. In previously reported cross-Ullmann reactions catalysed by a single metal, two or more equivalents of one reactant were required to produce yields that were consistently greater than 60%, and ratios closer to 1/1 usually resulted in dramatically lower yields[22]. However, in our multimetallic cross-Ullmann reaction, an excess of aryl triflate did not dramatically alter the yield of cross-product, although in some cases an excess of aryl bromide did improve our results. For example, three equivalents of 3-bromothiophene improved the cross-coupled yield from 64% to 85% (**18** in Fig. 3). However, an excess of aryl bromide did not improve yields with pyridine substrates.

The selectivity and rate of the reactions were also improved by the addition of potassium fluoride (KF). In all examples studied, the addition of one equivalent of KF accelerated the rate of the couplings, and, with the exception of reactions with pyridyl chlorides and activated aryl bromides, a higher selectivity for cross-product over dimerization was observed (Extended Data Fig. 2). In the coupling of bromobenzene with 4-methoxyphenyl triflate (Table 1, entry 1), 89% of substrate **2** was consumed at 24 hours, producing a 62% yield of product **3**. In the presence of KF, however, 91% conversion of **2** occurred in 11 hours with a 77% yield of **3** (Table 1, entry 2). The beneficial effect of KF seen here is consistent with observations of other cross-coupling reactions that are catalysed by nickel[23] and palladium[24,25]. In nickel catalysis, KF can reduce the amount of dimeric products formed; in palladium catalysis, KF can alter the oxidative addition selectivity for carbon–triflate bonds over carbon–X bonds through the formation of a palladium 'ate' complex. In our preliminary studies, it appears that both the potassium ion and the fluoride ion play a role in enhancing catalysis. Although optimization was performed in an inert-atmosphere glove box, reactions could be set up on the bench top as long as standard air-free techniques were used (Table 1, entry 13). Reactions run without added catalysts did not consume starting materials (entry 14), demonstrating that zinc itself is slow to react with the aryl electrophiles under these conditions and arguing against an intermediary role of arylzinc reagents in this cross-Ullman reaction. (For further details on reaction condition optimization, see Supplementary Information, Charts S4–S9).

Applying these reaction conditions to a variety of aryl electrophiles demonstrates the generality of this method (Fig. 3). Both

Table 1 | Conditions for the multimetal-catalysed cross-Ullmann reaction

Entry	w	x	y	z	3*	4*	5*
1†	5	5	5	5	77	7	4
2	5	5	5	5	62	21	14
3	10	10	10	10	70	17	10
4	0	0	10	10	3	4	1
5	0	10	10	0	0	15	0
6	10	10	0	0	6	42	6
7	10	0	0	10	1	1	0
8	0	5	10	5	4	7	1
9	10	5	0	5	1	16	0
10‡	10	10	10	10	67	13	12
11	5	5	10	10	61	20	14
12	10	10	5	5	66	18	13
13§	10	10	10	10	65	12	10
14	0	0	0	0	0	0	0

Reactions were run on a 0.5 mmol scale in 2 ml DMF.
*Yields were determined by gas chromatography and are expressed as area per cent, uncorrected. Remaining yields consist of unreacted starting material or hydrodehalogenation byproducts.
†One equivalent of KF was added.
‡Ligands and metal salts were not individually pre-stirred to ensure ligation. All solid reagents were added together, followed by solvent and liquids.
§Instead of being set up in a glove box, the reaction was set up on the bench top using standard inert atmosphere techniques.
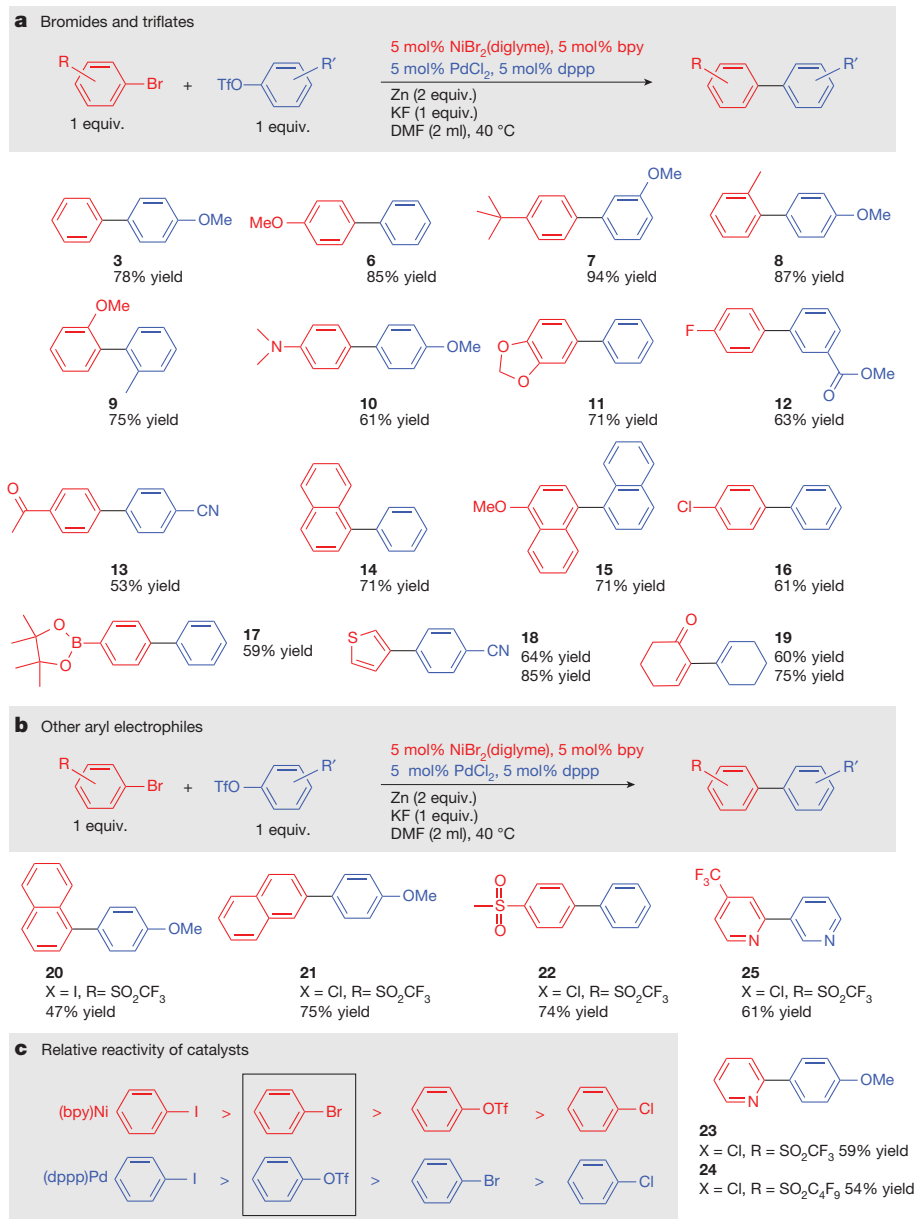
Figure 3 | Substrate scope. a, Reactions of aryl and vinyl bromides with aryl and vinyl triflates. **b**, Reactions of other aryl electrophiles. **c**, Relative reactivity of the palladium and nickel catalysts. To generate products **18**, **23**, **24** and **25**, no KF was used. For product **18**, yield was increased from 64% to 85% when three (rather than one) equivalents of 3-bromothiophene were used. For product **19**, yield was increased from 60% to 75% when two equivalents of 2-bromocyclohexenone were used. The reaction to produce product **3** was replicated independently by three different researchers in our group using the procedure given in the Supplementary Information.

electronically similar and electronically disparate aryl bromides and triflates couple with good yields, but the coupling of a highly activated substrate with a deactivated substrate results in a low yield. Aryl substrates with *ortho*-substitution can be coupled (Fig. 3a, **8** and **14**), and the synthesis of 2,2′-disubstituted biaryls and 2,2′-binaphthyls can also be achieved (Fig. 3a, **9** and **15**). The tolerance of this multimetallic method to different functional groups—including amine, ether, ester, aldehyde, sulfone, and acetal groups—is promising. In addition, boronic acid ester containing product **17** was formed without competing Suzuki coupling, despite the presence of palladium and KF, demonstrating complementarity with established cross-coupling methods. Although the combination of other halides and pseudohalides under these reaction conditions resulted in lower yields (see Supplementary Information, Charts S1–S3), we found several synthetically useful exceptions. While reactions with aryl iodides generally resulted in rapid formation of biaryl (see Supplementary Information, Chart S3; both palladium and nickel react with aryl iodides more quickly than with aryl triflates), naphthyl iodide could be coupled to form product **20** with 47% yield (Fig. 3b). Similarly, 2-naphthyl chloride could be coupled to form product **21** with 75% yield, suggesting that the nickel catalyst has a

strong affinity for naphthyl substrates that is not shared by the palladium catalyst[26]. The higher yield of the coupled product obtained when using naphthyl chloride rather than naphthyl iodide suggested the possibility of coupling other activated aryl chlorides with this system. Indeed, an aryl chloride with a sulfonyl group in the *para* position was successfully coupled with high selectivity to give product **22**. By contrast, reactions of unactivated aryl chlorides formed large amounts of symmetric biaryl from the aryl triflate. Bromothiophene and pyridyl halides and triflates coupled with reasonable yield, including the synthesis of 2,3′-bipyridine (product **25**). Finally, the chemistry could be extended to the synthesis of diene (product **19**) from α-bromocyclohexenone and cyclohexenyl triflate. To our knowledge, this is the first cross-Ullmann reaction to form a diene product.

Although we have not yet conducted a detailed study of the mechanism, current evidence is consistent with our original proposal (Fig. 1). We propose that, after the two arylmetal intermediates are formed by oxidative addition, a mechanism analogous to the 'persistent radical effect' allows for the observed cross-selectivity in transmetallation[27,28]. In the persistent radical effect, one radical is unreactive with itself and stable, while the other radical is highly reactive and

unselective. In our system, (dppp)Pd(Ar²) triflate is stable to further reaction with itself (Fig. 2d) and accumulates in solution. By contrast, (bpy)Ni(Ar¹) bromide is a transient intermediate that quickly reacts with itself (Fig. 2c) or with (dppp)Pd(Ar²) triflate (Fig. 2b). Selectivity arises from the relative concentrations of these two complexes in solution, similar to what occurs during multimetallic reactions with group 11 organometallic intermediates, which are generally slow to homocouple[11]. Consistent with our observations (Table 1), the selectivity of this cross-Ullmann reaction is predicted to be independent of the amount of palladium and nickel added. Furthermore, our proposed mechanism suggests that the formation of (bpy)Ni(Ar¹) bromide would limit the rate of product formation. Therefore, a higher concentration of nickel, but not of palladium, should result in a higher rate of product formation. We found this to be the case: reactions that were run with a higher concentration of nickel were faster (24% yield versus 7% yield at 4 hours for 25 mM versus 12.5 mM nickel), but the concentration of palladium did not affect the rate of product formation (11% yield versus 7% yield at 4 hours for 25 mM versus 12.5 mM palladium). We are now working to further illuminate the mechanism of this reaction and the origin of the cross-selectivity.

After a century of developing strategies for the formation of biaryls, a general method to selectively form unsymmetrical biaryls directly from two aryl electrophiles has been realized. In addition to the immediate utility of this transformation, the general principle of selective transmetallation between a persistent catalytic intermediate and a transient, reactive organometallic intermediate should spur the development of new multimetallic reactions beyond the cross-Ullman coupling, such as C–H arylation[29].
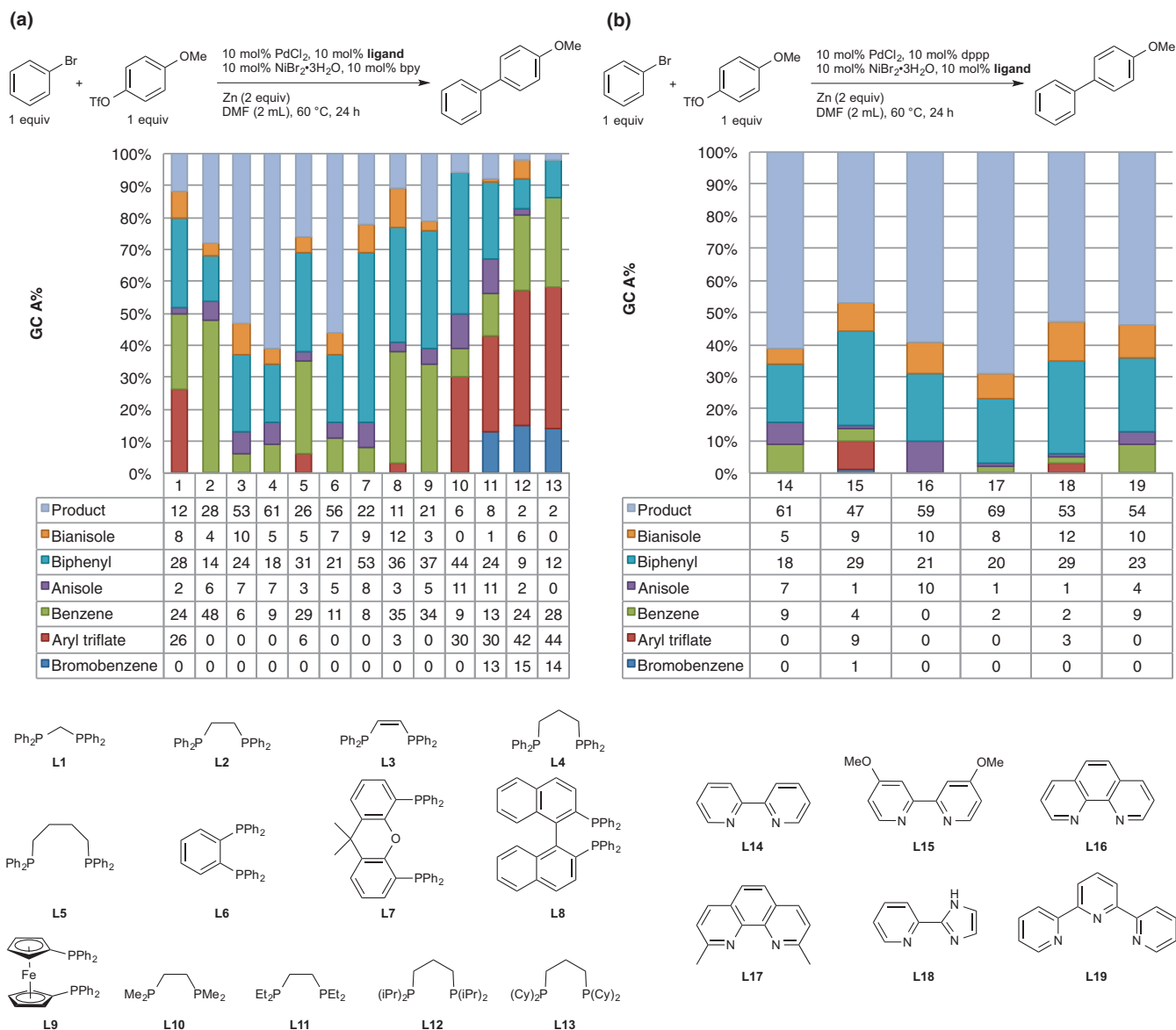
1. Magano, J. & Dunetz, J. R. Large-scale applications of transition metal-catalyzed couplings for the synthesis of pharmaceuticals. *Chem. Rev.* **111,** 2177–2250 (2011).
2. Busacca, C. A., Fandrick, D. R., Song, J. J. & Senanayake, C. H. The growing impact of catalysis in the pharmaceutical industry. *Adv. Synth. Catal.* **353,** 1825–1864 (2011).
3. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54,** 3451–3479 (2011).
4. Allen, A. E. & MacMillan, D. W. C. Synergistic catalysis: a powerful synthetic strategy for new reaction development. *Chem. Sci.* **3,** 633–658 (2012).
5. Shibasaki, M. & Yamamoto, Y. *Multimetallic Catalysis in Organic Synthesis* (Wiley, 2004).
6. Smidt, J. *et al.* Olefinoxydation mit Palladiumchlorid-Katalysatoren. *Angew. Chem.* **74,** 93–102 (1962).
7. Takacs, J. & Jiang, X. The Wacker reaction and related alkene oxidation reactions. *Curr. Org. Chem.* **7,** 369–396 (2003).
8. Michel, B. W., Steffens, L. D. & Sigman, M. S. The Wacker oxidation. *Org. React.* **84,** 75–414 (2014).
9. Sonogashira, K., Tohda, Y. & Hagihara, N. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedr. Lett.* **16,** 4467–4470 (1975).
10. Chinchilla, R. & Nájera, C. The Sonogashira reaction: a booming methodology in synthetic organic chemistry. *Chem. Rev.* **107,** 874–922 (2007).
11. Pérez-Temprano, M. H., Casares, J. A. & Espinet, P. Bimetallic catalysis using transition and group 11 metals: an emerging tool for C–C coupling and other reactions. *Chem. Eur. J.* **18,** 1864–1884 (2012).
12. Ko, S., Kang, B. & Chang, S. Cooperative catalysis by Ru and Pd for the direct coupling of a chelating aldehyde with iodoarenes or organostannanes. *Angew. Chem. Int. Edn* **44,** 455–457 (2005).
13. Ullmann, F. & Bielecki, J. Ueber Synthesen in der Biphenylreihe. *Chem. Ber.* **34,** 2174–2185 (1901).
14. Fanta, P. E. The Ullmann synthesis of biaryls. *Synthesis* **1974,** 9–21 (1974).
15. Hassan, J., Sévignon, M., Gozzi, C., Schulz, E. & Lemaire, M. Aryl-aryl bond formation one century after the discovery of the Ullmann reaction. *Chem. Rev.* **102,** 1359–1470 (2002).
16. Ackermann, L. *Modern Arylation Methods* (Wiley, 2009).
17. Stuart, D. R. & Fagnou, K. The catalytic cross-coupling of unactivated arenes. *Science* **316,** 1172–1175 (2007).
18. Weber, D. & Gagne, M. R. Pd(0)/Au(ɪ) redox incompatibilities as revealed by Pd-catalyzed homo-coupling of Arylgold(ɪ)-complexes. *Chem. Commun.* **47,** 5172–5174 (2011).
19. Kamikawa, T. & Hayashi, T. Control of reactive site in palladium-catalyzed Grignard cross-coupling of arenes containing both bromide and triflate. *Tetrahedr. Lett.* **38,** 7087–7090 (1997).
20. Everson, D. A., Jones, B. A. & Weix, D. J. Replacing conventional carbon nucleophiles with electrophiles: nickel-catalyzed reductive alkylation of aryl bromides and chlorides. *J. Am. Chem. Soc.* **134,** 6146–6159 (2012).
21. Panteleev, J., Zhang, L. & Lautens, M. Domino rhodium-catalyzed alkyne arylation/ palladium-catalyzed N arylation: a mechanistic investigation. *Angew. Chem. Int. Edn* **50,** 9089–9092 (2011).
22. Gomes, P., Fillon, H., Gosmini, C., Labbé, E. & Périchon, J. Synthesis of unsymmetrical biaryls by electroreductive cobalt-catalyzed cross-coupling of aryl halides. *Tetrahedron* **58,** 8417–8424 (2002).
23. Hatakeyama, T., Hashimoto, S., Ishizuka, K. & Nakamura, M. Highly selective biaryl cross-coupling reactions between aryl halides and aryl Grignard reagents: a new catalyst combination of N-heterocyclic carbenes and iron, cobalt, and nickel fluorides. *J. Am. Chem. Soc.* **131,** 11949–11963 (2009).
24. Littke, A. F., Dai, C. & Fu, G. C. Versatile catalysts for the Suzuki cross-coupling of arylboronic acids with aryl and vinyl halides and triflates under mild conditions. *J. Am. Chem. Soc.* **122,** 4020–4028 (2000).
25. Proutiere, F. & Schoenebeck, F. Solvent effect on palladium-catalyzed cross-coupling reactions and implications on the active catalytic species. *Angew. Chem. Int. Edn* **50,** 8192–8195 (2011).
26. Rosen, B. M. *et al.* Nickel-catalyzed cross-couplings involving carbon–oxygen bonds. *Chem. Rev.* **111,** 1346–1416 (2011).
27. Studer, A. The persistent radical effect in organic synthesis. *Chem. Eur. J.* **7,** 1159–1164 (2001).
28. Fischer, H. The persistent radical effect: a principle for selective radical reactions and living radical polymerizations. *Chem. Rev.* **101,** 3581–3610 (2001).
29. Durak, L. J. & Lewis, J. C. Iridium-promoted, palladium-catalyzed direct arylation of unactivated arenes. *Organometallics* **33,** 620–623 (2014).
30. Montgomery, J. in *Organometallics in Synthesis: Fourth Manual* (ed. Lipshutz, B. H.) 319–428 (John Wiley & Sons, 2013).

(a)

10 mol% PdCl$_2$, 10 mol% **ligand**
10 mol% NiBr$_2$·3H$_2$O, 10 mol% bpy
Zn (2 equiv)
DMF (2 mL), 60 °C, 24 h

	1	2	3	4	5	6	7	8	9	10	11	12	13
Product	12	28	53	61	26	56	22	11	21	6	8	2	2
Bianisole	8	4	10	5	5	7	9	12	3	0	1	6	0
Biphenyl	28	14	24	18	31	21	53	36	37	44	24	9	12
Anisole	2	6	7	7	3	5	8	3	5	11	11	2	0
Benzene	24	48	6	9	29	11	8	35	34	9	13	24	28
Aryl triflate	26	0	0	0	6	0	0	3	0	30	30	42	44
Bromobenzene	0	0	0	0	0	0	0	0	0	0	13	15	14

(b)

10 mol% PdCl$_2$, 10 mol% dppp
10 mol% NiBr$_2$·3H$_2$O, 10 mol% **ligand**
Zn (2 equiv)
DMF (2 mL), 60 °C, 24 h

	14	15	16	17	18	19
Product	61	47	59	69	53	54
Bianisole	5	9	10	8	12	10
Biphenyl	18	29	21	20	29	23
Anisole	7	1	10	1	1	4
Benzene	9	4	0	2	2	9
Aryl triflate	0	9	0	0	3	0
Bromobenzene	0	1	0	0	0	0

L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, L17, L18, L19

Extended Data Figure 1 | Varying ligands on the nickel and palladium catalysts. Several different phosphine ligands for palladium (**a**) and amine ligands for nickel (**b**) were investigated. Although selectivity and yield of cross-product were sensitive to the identity of the phosphine ligand (**a**), a variety of different amine ligands were effective (**b**). Reactions conducted with **L15**, **L17**, and **L18** used 5 mol% catalyst loadings, were heated to 40 °C, and were monitored for 64 h. GC, gas chromatography; area per cent, uncorrected.

(a)

	0.0 equiv	0.20 equiv	0.5 equiv	1.0 equiv	2.0 equiv	3.0 equiv
▪ Product	62	37	79	77	79	72
▪ Bianisole	14	1	5	4	9	12
▪ Biphenyl	21	13	12	7	10	13
▪ Anisole	1	0	0	1	0	0
▪ Benzene	3	2	3	2	2	2
▪ Aryl triflate	0	28	1	9	0	0
▪ Bromobenzene	0	20	0	0	0	0

(c)

	No Salt	KF	LiF	CsF	KI	KBr	KOAc
▪ Product	62	79	66	27	35	34	46
▪ Bianisole	14	8	10	19	28	26	24
▪ Biphenyl	21	9	15	9	37	36	29
▪ Anisole	1	0	0	5	0	1	0
▪ Benzene	3	3	4	30	1	1	1
▪ Aryl triflate	0	0	3	11	0	2	0
▪ Bromobenzene	0	0	0	0	0	0	0

(b)

Equiv KF	t (h)	PhBr Remaining (% yield)	ArOTf Remaining (% yield)
0	11	30	38
	24	5	10
0.2	11	35	36
	24	20	28
0.5	11	13	15
	24	0	1
1	11	0	9
	24	0	0
2	11	1	0
	24	0	0
3	11	0	0
	24	0	0

(d)

Additive	t (h)	PhBr Remaining (% yield)	ArOTf Remaining (% yield)
No Salt	3	42	47
	24	5	10
KF	3	30	31
	24	0	0
LiF	3	39	39
	24	0	3
CsF	3	3	37
	24	0	11
KI	3	5	37
	24	0	0
KBr	3	0	23
	24	0	2
KOAc	3	2	26
	24	0	0

Extended Data Figure 2 | The effect of potassium and fluoride salts on the selectivity of the multimetal-catalysed cross-Ullmann reaction. The presence of potassium fluoride (KF) in cross-couplings has been demonstrated to improve the product yield and rate of reactions while minimizing the formation of dimeric products. This 'KF effect' can be attributed to a variety of different factors, including the formation of 'ate' complexes, the removal of a halide from a metal complex, or the formation of a fluoride-bridged complex, which could facilitate an oxidative addition or transmetallation step[23,25]. **a, b**, To investigate whether KF could be beneficial for the nickel and palladium multimetallic system, we included various amounts of this additive under standard reaction conditions. The resulting selectivity and rate data are shown. **c, d**, Once the advantageous role of KF was confirmed, other potassium and fluoride salts were tested. The resulting selectivity and rate data are shown. The presence of potassium resulted in faster reaction rates, while the presence of fluoride reduced the amount of bianisole and biphenyl byproducts, improving the yield of cross-product. (Reactions without additives took 32 h to complete.)

LETTER

Western US intermountain seismicity caused by changes in upper mantle flow

Thorsten W. Becker[1], Anthony R. Lowry[2], Claudio Faccenna[3], Brandon Schmandt[4], Adrian Borsa[5] & Chunquan Yu[6]

Understanding the causes of intraplate earthquakes is challenging, as it requires extending plate tectonic theory to the dynamics of continental deformation. Seismicity in the western United States away from the plate boundary is clustered along a meandering, north–south trending 'intermountain' belt[1]. This zone coincides with a transition from thin, actively deforming to thicker, less tectonically active crust and lithosphere. Although such structural gradients have been invoked to explain seismicity localization[2,3], the underlying cause of seismicity remains unclear. Here we show results from improved mantle flow models that reveal a relationship between seismicity and the rate change of 'dynamic topography' (that is, vertical normal stress from mantle flow). The associated predictive skill is greater than that of any of the other forcings we examined. We suggest that active mantle flow is a major contributor to seismogenic intraplate deformation, while gravitational potential energy variations have a minor role. Seismicity localization should occur where convective changes in vertical normal stress are modulated by lithospheric strength heterogeneities. Our results on deformation processes appear consistent with findings from other mobile belts[4], and imply that mantle flow plays a significant and quantifiable part in shaping topography, tectonics, and seismic hazard within intraplate settings.

Interplate earthquakes are well described by plate tectonics as a consequence of horizontal motion gradients within the surface boundary layer of thermo-chemical mantle convection. For seismicity away from plate boundaries, mantle convection is the fundamental cause as well. However, it is unclear to what extent the location and rate of seismicity is determined by the present-day forcing from active mantle flow, rather than its integrated effect on the buoyancy and rheology of continental lithosphere over geological history. To address this question, we focus on intraplate seismicity in the western US mobile belt, where seismicity and crustal deformation of predominantly extensional style is found around the Snake River plain and Yellowstone hotspot, and following the boundary between the eastern Basin and Range province and the Colorado Plateau further to the south (Fig. 1a–c).

From a lithospheric perspective, lateral variations in density are also expected to have a role in causing seismic strain release. These density variations will lead to lateral variations in gravitational potential energy (GPE; Fig. 1d), that is, the depth-integrated density moment. GPE variations are associated with deviatoric stresses[5] and have been linked with tectonic deformation in the region[6]. Seismicity may further localize at lithospheric strength boundaries, arising from changes in crustal thickness, lithospheric composition, or geotherms[2,3,7].

Lithospheric heterogeneities evolve over long timescales from mass transport processes associated with plate tectonics and sublithospheric dynamics, including small-scale mantle convection and lithospheric instabilities[7–9]. Present-day active mantle flow may also play a part, and a link between large-scale downwellings and intraplate seismicity was suggested for the 1811/1812 New Madrid ($M \approx 7$) earthquakes, for example[10]. Advances in structural seismology[3,11,12] now allow calculations of mantle flow on the scales appropriate for our study region[13]. Such models appear to capture deep mantle mass transport[14], and provide an explanation for anomalous, non-isostatic topography[13,15] (Fig. 1e). The causes for the general state of stress in this part of the North American plate have been explored in numerous studies[16–19], and geodetically imaged strain-rate patterns in the region have been linked to small-scale mantle flow[8]. However, to our knowledge, no prior geodynamic effort has attempted to predict the spatial distribution of seismicity.

A visual inspection shows an association between seismicity (Fig. 1a, b) and GPE (Fig. 1d), suggesting that lateral gradients in structure are relevant to where earthquakes are found. It is also striking how the rate change of predicted dynamic topography appears to coincide with the regions of intermountain seismicity (Fig. 1f). To quantify these relationships, we computed Molchan error curves[20] (Fig. 2a) and associated skill, S (Fig. 2b), a statistical metric developed to measure the predictive power of earthquake forecasts[21]. As defined here, $-0.5 \leq S \leq 0.5$, with $S = 0$ for a random forecast and $S = 0.5$ for a perfect, positive correlation. Actual model skill can be compared with the best case, the seismicity function (Fig. 1b), for which $S = 0.42$ because of smoothing (Fig. 2a).

Figure 2b shows that lithospheric structure models such as Moho depth and effective elastic thickness, an integral measure of strength, are poor predictors of seismicity, while the two GPE models show significantly positive skill. This implies that seismicity is preferentially found in regions of elevated GPE where an extensional state of stress should be found. When gradients of structural models are considered (Fig. 2b), Moho gradients have small but significantly positive skill, as expected[3,7]. Variations in elastic thickness and GPE model gradients show poor skill.

If we consider mantle flow model predictions, the shear stress amplitude within the viscous flow model's lithosphere is a very poor predictor (Fig. 2b). The vertical normal stress from present-day mantle flow, expressed here as an equivalent 'dynamic topography'[22] that would be generated in the absence of lithospheric strength and dynamics (Fig. 1e), has a lightly positive S; that is, however, not outside the 95% confidence range. In contrast, the inferred rate of change of dynamic topography (Fig. 1f) shows a strong, positive skill of $S = 0.26$. This skill is larger than for any of the other models, and is only surpassed by the match of seismicity with the maximum geodetic shear strain rates (Fig. 1c), where S is of comparable magnitude, $S = 0.28$, and Molchan curves have a similar functional form (Fig. 2a). The strong relationship of seismicity to geodetic strain rates is of interest, since geodetic and seismic strain release need not be aligned perfectly throughout the seismic cycle, and aseismic deformation could be decorrelated, which (apparently) it is not.

However, crustal deformation forecast skill cannot answer the question about the actual cause of deformation, and only serves as a point of

[1]Department of Earth Sciences, University of Southern California, Los Angeles, California 90089–0740, USA. [2]Department of Geology, Utah State University, Logan, Utah 84322–4505, USA. [3]Department of Sciences, Università Roma Tre, 00146 Roma, Italy. [4]Department of Earth and Planetary Sciences, The University of New Mexico, Albuquerque, New Mexico 87131, USA. [5]Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093–0225, USA. [6]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
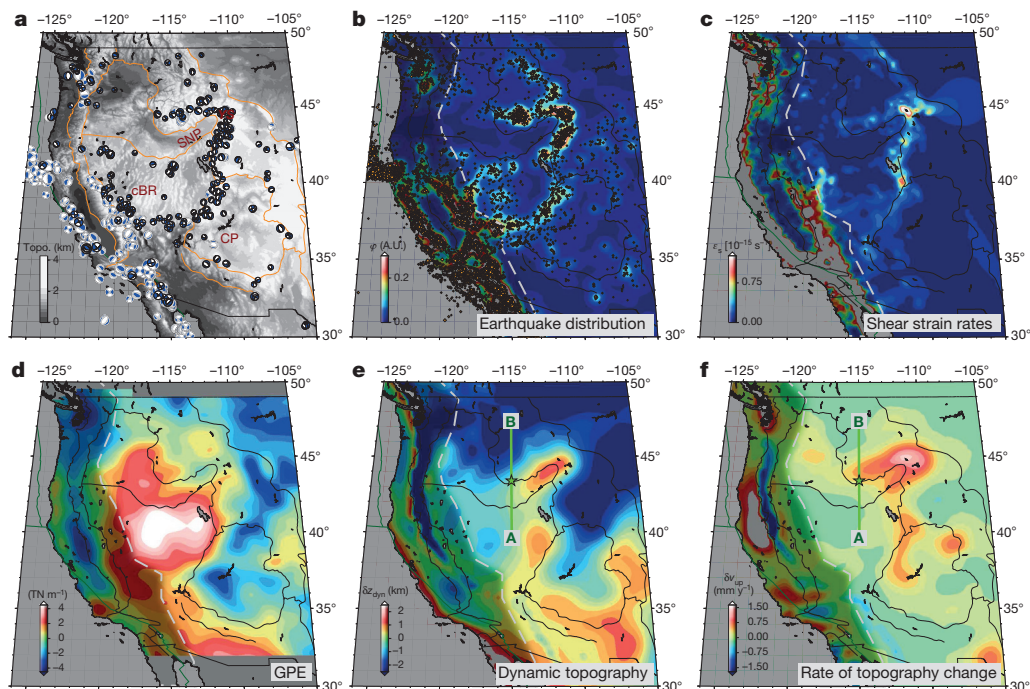
Figure 1 | Structural and geodynamic maps for the study region.
a, Topography (Topo.), topographic provinces (orange lines) and focal mechanisms (grey and black from the global Centroid-Moment-Tensor (gCMT)[29] and Saint Louis University[1] catalogues, respectively (gCMT catalogue accessed in December 2014)). cBR, central Basin and Range; CP, Colorado Plateau; SNP, Snake River Plain. **b**, Shallow (depths ≤50 km) seismicity and density function, ϕ. A.U., arbitrary units. **c**, Geodetic shear strain rates from GPS[30]. **d**, GPE estimate based on our preferred crustal model[13]. **e**, Dynamic topography inferred from mantle flow[13]. **f**, Rate change of dynamic topography. Shading marked by dashed lines indicates plate boundary regions excluded from most analysis, and **e**, **f**, show the profile used in Fig. 3. See Methods for details.

comparison. Normal geodetic strain rates (Extended Data Fig. 1e) also show positive skill (Fig. 2b), consistent with most events being of extensional character (Fig. 1a). Such a signal in the crust might be expected if an active mantle upwelling (Figs 1f and 3) were to push up and locally extend the lithosphere. (Additional structural models and extension of Molchan analysis to plate boundary seismicity are shown in Extended Data Figs 1–3, and compared with total and wavelength-dependent correlation in Extended Data Figs 4 and 5.)

Inspection of Fig. 1e, f reveals that regions of highest rate of change in dynamic topography correspond to gradients in dynamic topography. The physics of this process can be illustrated by examining mantle flow model profiles across a high seismicity band (Fig. 3). Depending on their buoyancy, upper mantle anomalies produce

upwellings or downwellings (and positive or negative dynamic topography), respectively. However, both types of anomalies can cause positive rate change of dynamic topography, since the upward push of a rising, positive anomaly increases dynamic topography, and the sinking of a negative anomaly reduces negative dynamic topography over time[23,24] (Extended Data Fig. 6). This means that the constructive superposition of anomalies from paired upwellings and downwellings focuses the signal in the intermediate region where most seismicity lies. From Fig. 1e, it appears that mantle upwellings are the main control on the spatial distribution of seismicity in our study region[13,15]. However, the mechanism of change of vertical stressing from mantle flow also applies above downwellings, broadly consistent with earlier work on the New Madrid events[10], albeit on different spatial scales.
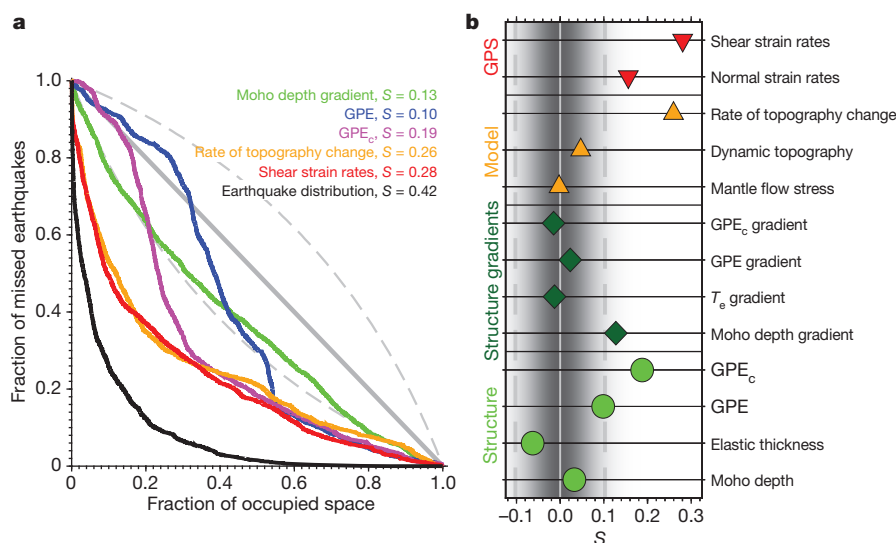

Figure 2 | Analysis of predictive power for intraplate seismicity. **a**, Molchan curves[20], where the diagonal indicates a random prediction, and curves outside the dashed lines are ≳95% significant. **b**, Skill S (that is, the area above the Molchan curve minus 0.5), where dashed lines indicate the regions outside of which results are 95% significant. We show results from structural models (Moho depth[11], effective elastic thickness T_e (ref. 11), and two kinds of GPE models (regular (Fig. 1d) and compensated, GPE$_c$), gradients thereof, geodynamic model predictions (compare with Fig. 1e, f) 'mantle flow stress', that is, the second invariant of the deviatoric stress), and geodetic models[30] (compare with Fig. 1c). See Methods for details.
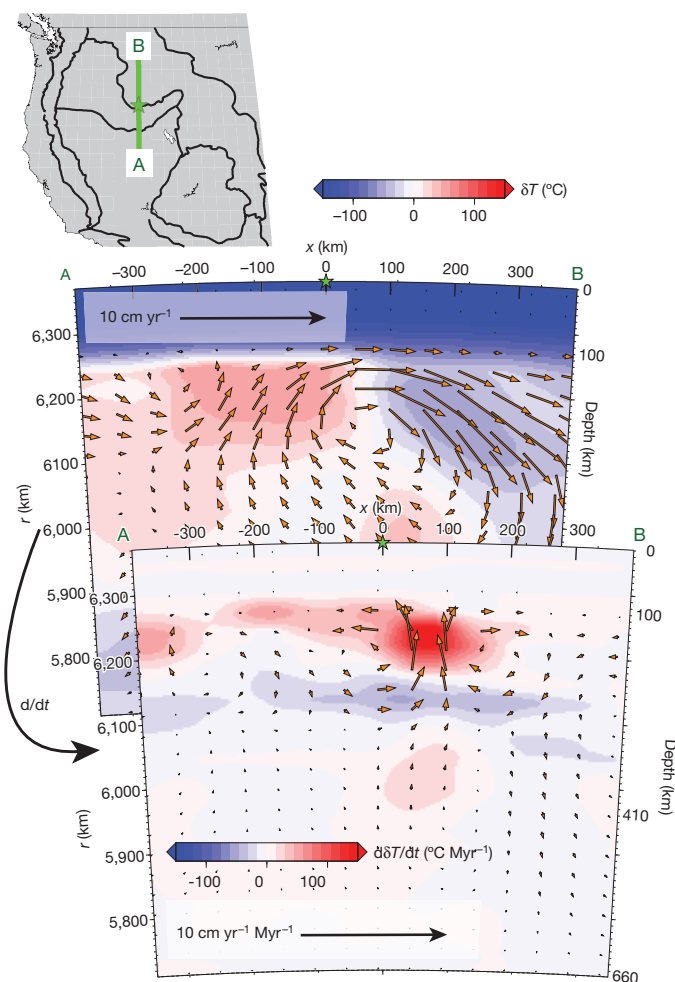
Figure 3 | Illustration of mantle flow forcing in the study region. Inferred temperature anomalies and mantle flow velocities (background) and temporal change thereof (foreground) are shown along a north–south profile. Profile location, corresponding dynamic topography, and rate change thereof are shown in Fig. 1e and f, respectively. Also see Extended Data Fig. 6.

Figures 1–3 suggest that the change in vertical normal stresses from mantle convection, coupled with modulation from lithospheric structure, might explain intermountain seismicity. Any such superposition will probably be nonlinear and wavelength dependent in a rheologically complex, visco-elasto-plastically deforming lithosphere. Results from a simplified, linear, multivariable regression imply that both GPE and structural gradients contribute positively, although correlations are improved only moderately relative to that of vertical stressing rates alone (Extended Data Fig. 7). Nonetheless, the physical interpretation of the combined model is appealing: intraplate seismicity responds to changes in vertical stressing rates from mantle flow, and is found in regions with elevated GPE. Moho depth gradients and/or variations in lithospheric thickness then contribute to the localization of seismicity, presumably since zones of lithospheric and crustal weakness lead to locally enhanced strain rates.

Molchan skill or correlation are imperfect metrics of agreement, and even large skill does not imply causality. However, our analysis suggests that intraplate seismicity may be determined by the interplay of active mantle upwellings and lithospheric heterogeneity, reconciling 'deep'[8,10,13] and 'shallow'[2,3,6] points of view on the origin of crustal deformation. GPE gradients themselves appear to play a minor part in driving seismicity in this region; rather, high GPE zones seem to mark domains that are buoyed by deep mantle flow[15] and so primed for extensional failure.

A preliminary analysis of geodetically determined vertical motions from Global Positioning System (GPS) data indicates that patterns of surface uplift do not correlate with our predictions of rate change of dynamic topography, except at some wavelengths (Extended Data Fig. 8). This lack of correlation partly stems from neglect of lithospheric strength by the viscous flow model, which would greatly dampen uplift on scales $\lesssim 200 \dots 600$ km in the region[15]. Effects such as the visco-elastic response to brittle deformation may also contribute strongly to the observed uplift. It is clear that fully resolving the mechanics of intraplate seismicity and topography has to await an integrated mechanical model that can explain a range of data from different disciplines—including geologically constrained uplift—and that improves on rheology and compositional realism. However, improvements in imaging and modelling capabilities, such as those facilitated by EarthScope (http://www.earthscope.org), have led to geodynamic models that can already be queried on the relevant scales.

Scenarios involving vertical mantle flow loading and intraplate seismicity may be of global relevance and apply in other mobile belts, such as the Mediterranean–Middle East region, where deformation is diffuse and perhaps related to small-scale convection[4,25]. There, the effect of mantle forcing may also be relevant along complex plate boundaries. An example are the Apennines, where surface deformation appears to be related to changes in dynamic topography[26], leading to elevated GPE[27], like in the western United States. This highlights the importance of ongoing mantle convection on sub-plate scales of hundreds to thousands of kilometres, leading to intraplate deformation and seismic activity.

Our analysis suggests that the mantle dynamical perturbation is most important to the deformation behaviour of the crust. Intraplate seismicity may require a disequilibrium perturbation to a system that is already in a critical state of equilibrium stress. This provides a hypothesis for the cause of continental deformation that is quantitatively testable in an increasing number of geographic settings.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Herrmann, R. B., Benz, H. & Ammon, C. J. Monitoring the earthquake process in North America. *Bull. Seismol. Soc. Am.* **101,** 2609–2625 (2011).
2. Lowry, A. R. & Smith, R. B. Strength and rheology of the western U.S. Cordillera. *J. Geophys. Res.* **100,** 17947–17963 (1995).
3. Levander, A. & Miller, M. S. Evolutionary aspects of the lithosphere discontinuity structure in the western U.S. *Geochem. Geophys. Geosys.* **13,** 1525–2027 (2012).
4. Faccenna, C. & Becker, T. W. Shaping mobile belts by small-scale convection. *Nature* **465,** 602–605 (2010).
5. Artyushkov, E. V. Stresses in the lithosphere caused by crustal thickness inhomogeneities. *J. Geophys. Res.* **78,** 7675–7708 (1973).
6. Jones, C. H., Unruh, J. R. & Sonder, L. J. The role of gravitational potential energy in active deformation in the southwestern United States. *Nature* **381,** 37–41 (1996).
7. Bird, P. Lateral extrusion of lower crust from under high topography, in the isostatic limit. *J. Geophys. Res.* **96,** 10275–10286 (1991).
8. Fay, N. P., Bennett, R. A., Spinler, J. C. & Humphreys, E. D. Small-scale upper mantle convection and crustal dynamics in southern California. *Geochem. Geophys. Geosys.* **9,** Q08006 (2008).
9. van Wijk, J. W. *et al.* Small-scale convection at the edge of the Colorado Plateau: Implications for topography, magmatism, and evolution of Proterozoic lithosphere. *Geology* **38,** 611–614 (2010).
10. Forte, A. M., Mitrovica, J. X., Moucha, R., Simmons, N. A. & Grand, S. P. Descent of the ancient Farallon slab drives localized mantle flow below the New Madrid seismic zone. *Geophys. Res. Lett.* **34,** L04308 (2007).
11. Lowry, A. R. & Pérez-Gussinyé, M. The role of crustal quartz in controlling Cordilleran deformation. *Nature* **471,** 353–357 (2011).
12. Schmandt, B. & Lin, F.-C. P and S wave tomography of the mantle beneath the United States. *Geophys. Res. Lett.* **41,** 6342–6349 (2014).
13. Becker, T. W., Faccenna, C., Humphreys, E. D., Lowry, A. R. & Miller, M. S. Static and dynamic support of western U.S. topography. *Earth Planet. Sci. Lett.* **402,** 234–246 (2014).
14. Schmandt, B., Jacobsen, S. D., Becker, T. W., Liu, Z. & Dueker, K. G. Dehydration melting at the top of the lower mantle. *Science* **344,** 1265–1268 (2014).

15. Lowry, A. R., Ribe, N. M. & Smith, R. B. Dynamic elevation of the Cordillera, western United States. *J. Geophys. Res.* **105,** 23371–23390 (2000).
16. Flesch, L. M., Holt, W. E., Haines, A. J. & Shen-Tu, B. Dynamics of the Pacific–North American plate boundary in the western United States. *Science* **287,** 834–836 (2000).
17. Humphreys, E. D. & Coblentz, D. North American dynamics and western U.S. tectonics. *Rev. Geophys.* **45,** RG3001 (2007).
18. Forte, A. M., Moucha, R., Simmons, N., Grand, S. & Mitrovica, J. Deep-mantle contributions to the surface dynamics of the North American continent. *Tectonophys.* **481,** 3–15 (2010).
19. Ghosh, A., Becker, T. W. & Humphreys, E. D. Dynamics of the North American continent. *Geophys. J. Int.* **194,** 651–669 (2013).
20. Molchan, G. M. & Kagan, Y. Y. Earthquake prediction and its optimization. *J. Geophys. Res.* **97,** 4823–4838 (1992).
21. Zechar, J. & Jordan, T. Testing alarm-based earthquake predictions. *Geophys. J. Int.* **172,** 715–724 (2008).
22. Richards, M. & Hager, B. H. Geoid anomalies in a dynamic Earth. *J. Geophys. Res.* **89,** 5987–6002 (1984).
23. Gurnis, M., Mitrovica, J. X., Ritsema, J. & van Heijst, H.-J. Constraining mantle density structure using geological evidence of surface uplift rates: the case of the African superplume. *Geochem. Geophys. Geosys.* **1,** 1020 (2000).
24. Conrad, C. P. & Husson, L. Influence of dynamic topography on sea level and its rate of change. *Lithosphere* **1,** 110–120 (2009).
25. Faccenna, C. *et al.* Mantle dynamics in the Mediterranean. *Rev. Geophys.* **52,** 283–332 (2014).
26. Faccenna, C., Becker, T. W., Miller, M. S., Serpelloni, E. & Willett, S. D. Isostasy, dynamic topography, and the elevation of the Apennines of Italy. *Earth Planet. Sci. Lett.* **407,** 163–174 (2014).
27. D'Agostino, N. & McKenzie, D. Convective support of long wavelength topography in the Apennines (Italy). *Terra Nova* **11,** 234–238 (1999).
28. Wessel, P. & Smith, W. H. F. New, improved version of the Generic Mapping Tools released. *Eos Trans. AGU* **79,** 579 (1998).
29. Ekström, G., Nettles, M. & Dziewoński, A. M. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Inter.* **200–201,** 1–9 (2012).
30. Kreemer, C., Blewitt, G. & Klein, E. C. A geodetic plate motion and Global Strain Rate Model. *Geochem. Geophys. Geosys.* **15,** 3849–3889 (2014).

Author Contributions T.W.B. conducted the analysis and geodynamic modelling and wrote the paper with A.R.L. T.W.B., A.R.L. and C.F. designed the analysis and contributed to the writing. B.S. and C.Y. advised on geophysical constraints, A.B. provided assistance with GPS data, and all authors collaborated on interpretation of results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.W.B. (twb@usc.edu).

METHODS

Lithospheric structure models. To address how much of the seismicity patterns (Fig. 1b) can be explained by lithospheric structure and gradients thereof, we mainly consider a Moho depth map and estimates of effective elastic thickness, T_e, from ref. 11; other Moho maps lead to similar results (Extended Data Fig. 2). Values of T_e should have some relation to the integrated lithospheric strength, but the driving forces for deformation within the lithosphere should instead be related to the density moment, gravitational potential energy (GPE). We therefore construct a GPE model, integrating our reference crustal model[13] down to an arbitrary compensation depth of $H = 150$ km. This model is based on crustal density and Moho estimates from ref. 11, and a homogeneous mantle lithosphere. Other crustal models will look similar in terms of GPE[13].

As expected, the resulting GPE model (Fig. 1d) shares similarities with the corresponding residual (non-Airy) topography[13], and the main feature away from the plate boundary is the anomalously high GPE within the Basin and Range[6,16]. The step towards relatively low GPE towards the Colorado Plateau corresponds to an increase in crustal thickness towards the east. The broad agreement of this gradient with seismicity patterns motivated the suggestion that lithospheric structure gradients relate to seismicity[2,3]. However, note that in the regions north of the Snake River Plain and Yellowstone, for example, Moho depth gradients are not well aligned with seismicity.

The crustal density model that underpins the GPE model of Fig. 1d is not in isostatic equilibrium, a point that can be interpreted as indicating predominant mantle flow support[13], or lithospheric compensation[31]. Thus, we also consider a compensated GPE model, GPE_c, which is constructed by enforcing isostatic balance at H by allowing arbitrary density anomalies within the mantle lithosphere. GPE_c has anomaly amplitudes that are subdued by a factor of ~2 compared to GPE and looks very similar to other estimates[31], with the largest anomalies found in the Basin and Range as for GPE.

From all structural models, we also compute the absolute amplitude of the spatial gradient vector. We do so after smoothing each input field with a six standard deviation, $6\sigma = 100$ km width, Gaussian smoothing kernel.

Geodynamic models. Mantle convection models assume a purely viscous, Newtonian rheology and incompressible Stokes flow in the Boussinesq approximation. In line with earlier studies[23,24,32], the surface boundary conditions are free-slip for computations of dynamic topography, or prescribed plate motions in a North America fixed reference frame, if change in dynamic topography is estimated. This accounts for the effects of relative transport of anomalies by large-scale mantle flow, which only plays a minor part for our study.

The computations here are based on a regional sector of a spherical annulus with horizontal dimensions of ~7,000 km × ~7,000 km and 1,200 km depth, computed with a finite element method[33,34], at half the mesh size (~10 km), but else as in our earlier global models[13,19]. The box size ensures that all results shown here from the centre of the domain are not affected by boundary effects. Dynamic topography results (Fig. 1e), for example, are consistent with those from global computations[13]. Results are also stable with respect to further mesh refinement, and resolution is mainly limited by the structural information from the seismological input models rather than by the numerical methods.

We generally infer mantle temperature (or density) anomalies from the most recent, high-resolution, shear wave, v_S, tomography model of ref. 12. We first embed this regional model into a global tomographic model[35] to avoid edge effects, and then scale to density ρ, with a constant conversion factor of $d\ln\rho/d\ln v_S = 0.2$ below 100 km, for simplicity, as in earlier work[13,14,19]. Scaled temperatures and velocities arise from a Rayleigh number of 2.3×10^7. With a reference viscosity of $\eta_0 = 10^{21}$ Pa s, the lithosphere (0...100 km), asthenosphere (100...300 km), upper mantle (300...660 km), and lower mantle have viscosities of $50\eta_0$, $0.1\eta_0$, η_0 and $50\eta_0$, respectively.

For the new rate change of dynamic topography computations, we add a half-space cooling model for the oceanic plate and include lateral lithospheric thickness variations based on inferred LAB thickness[3]. These modifications are intended to improve realism close to the plate boundary; without the modifications, plate boundary proximal regions such as the Cascades show less negative change in dynamic topography (Fig. 1f). However, the modifications have a minor role in shaping rates of change of dynamic topography within the interior plate region, on which our seismicity analysis is focused. Apart from those modifications, the mantle flow models are identical in terms of the approach to our earlier, global mantle flow computations[13,14], explored with modifications such as anisotropic viscosities in ref. 19.

On the basis of the flow computations, we compute several stress quantities. Stress amplitudes for correlation analysis are obtained from the second invariant of the deviatoric stress tensor at ~10 km depth. The instantaneous 'dynamic topography' δz_{dyn} (Fig. 1e) is computed from the inferred deflections of the surface due to the vertical normal tractions exerted by present-day flow on the top of the

computational domain. Our estimates of dynamic topography here are very similar to those from rheologically simpler models[13], as expected, since assumptions about density anomalies within the top ~350 km are most important for dynamic topography on these scales, and dynamic topography amplitudes scale with density anomalies, to first order.

We also compute the temporal change of δz_{dyn}, that is, change in mantle-flow-induced vertical normal stress, $\delta\dot{z}_{dyn}$, by taking a finite, forward time derivative between two estimates of δz_{dyn} after accounting for advection and diffusion of mantle temperature anomalies as well as plate motions for ~10,000 years[23,24]. The physics of how different anomalies are expressed as δz_{dyn} and $\delta\dot{z}_{dyn}$ are illustrated in Extended Data Fig. 6 for synthetic anomalies and Fig. 3 for the study region.

Rate change of dynamic topography scales roughly with the density anomalies squared and inversely with asthenospheric viscosity[23], and those values, their lateral variations, and the general form of the appropriate constitutive relationship, are poorly known. Patterns of $\delta\dot{z}_{dyn}$ from flow are therefore much more stable than rates, for a given mantle density distribution, which is why we do not use any absolute amplitude information. The actual geological expression of change in dynamic topography as vertical rock motions will, moreover, be filtered by a range of processes such as erosion, magmatic intrusion, and tectonic shortening or rifting in a rheologically complex lithosphere[36,37], none of which are considered here.

Predictions from such mantle flow models are expected to be most strongly controlled by the input tomography model. Related differences can be illustrated by considering older models (Extended Data Fig. 9). These lead to flow estimates, and hence conclusions for skill, that are quantitatively different, but qualitatively similar to, those based on our reference model, ref. 12. This is expected because recent, regional tomographic models for the western United States agree in terms of broad velocity anomaly patterns[38]. Moreover, skill is near identical for flow based on the predecessor model for ref. 12, ref. 39 (Extended Data Fig. 9), which was found to best predict seismic delay times[40]. We expect its successor[12] to be even better constrained given that this model was able to rely on a wider aperture data set from EarthScope's USArray.

Another important issue is what assumptions are made regarding the existence of compositional anomalies[13,18,41,42], which are ignored here besides exclusion of the top 100 km of the lithosphere. This approach is motivated by an encouraging match between dynamic topography from our earlier flow computations and residual surface topography[13] and the overall ability of the flow computations to match large-scale plate motions, the geoid, and deep mass transport[14,19]. We therefore expect the broad patterns of Fig. 1e, f, and the conclusions based on Fig. 2, to be representative of the general expression of mantle flow. However, details of the model predictions may well be subject to further refinement if compositional and partial melt effects on seismic wave speeds, and hence the scaling to temperature, are better constrained.

Seismicity and Molchan predictive skill. We base all comparisons of different forward models with seismicity on a merger of all $M \geq 3$ events in two earthquake catalogues[43,44] that are shallower than 50 km. Our main evaluation of predictive power is based on constructing forecast error curves, typically attributed to Molchan[20]. Such curves can be computed for any scalar field at each instance in time, or for time-independent forecasts, as is done here. The curves are constructed by scanning through the range of represented scalar values of the forecast and plotting, for each value, the fractional number of earthquakes, y, inside regions where the scalar field is above the current value against the fractional area where scalars are below, x (Fig. 2a).

A random forecast will plot along the diagonal from $\{x, y\} = \{0, 1\}$ to $\{1, 0\}$, and deviations thereof, for example as measured by the area above the Molchan curve, can be used to quantify predictive power[21]. Here, we define the skill, S, as that area minus 0.5, such that $S = 0$ indicates purely random, and the extreme values of $S = 0.5$ and $S = -0.5$ perfect correlation and anti-correlation, respectively, in analogy to correlation metrics.

All fields are first smoothed with a $6\sigma = 100$ km width spatial Gaussian filter before constructing the Molchan curves. We explored other filters as well; while absolute S values are affected, relative performance of models was found to be similar. We estimate significance from 10,000 Monte Carlo simulations using spatial fields of Gaussian distributed (white noise) random numbers, smoothed in the same way as the actual scalar fields. The corresponding random S values are found to be approximately Gaussian distributed[21] (grey shading in Fig. 2 and Extended Data Fig. 2), and the ~95% significance range for our problem can be found outside $\pm 2\Sigma_S$, with standard deviation of $\Sigma_S \approx 0.05$.

We consider only seismicity away from the plate boundary for the Molchan curves in Fig. 2 and Extended Data Fig. 2, but Extended Data Fig. 3 shows results including all seismicity of Fig. 1b. For geodetically determined shear strain rates, the skill is larger when plate boundary seismicity is included, increased from $S = 0.28$ to $S = 0.35$. In contrast, S for the rate change of dynamic topography is

reduced from 0.26 to 0.19 when including the plate boundary. However, this S value is still larger than for the other models, such as Moho depth gradients (Extended Data Fig. 3). Moreover, the compensated GPE model, GPE_C, which indicated high skill in Fig. 2 for the intraplate events, loses predictive power when all events are considered: S is reduced from 0.19 to -0.06. These findings imply that rate change of dynamic topography may also contribute as a relevant forcing in interplate settings.

Smoothed seismicity and correlations. We also compute a seismicity field to estimate a background 'forecast' for the Molchan curves, and to use for additional correlation analysis. For this, we sum seismic moments on a $0.1° \times 0.1°$ grid, take the $\log_{10}$ of this sum, and then smooth this field with a $6\sigma = 150$ km width Gaussian filter to define the seismicity density function, ϕ, shown in arbitrary units in Fig. 1b. The skill for this smoothed seismicity field is $S = 0.42$, and provides a realistic upper bound for positive correlation forecasts for all scalar fields considered (Fig. 2a and Extended Data Fig. 2).

As for any such study, 'smoothed seismicity' maps, like those used for probabilistic forecasting[21], require choices that influence the scalar fields. Questions arise as to whether to sum magnitudes, number of events, moments, or Benioff strain, and how event kernels should be applied. The Molchan curves are based on individual events and do not suffer from such smoothing assumptions. However, both Molchan curves and smoothed seismicity are affected by earthquake catalogue completeness, which is limited by the duration of the instrumental record, and there will be temporal and spatial changes in completeness. We therefore tried a number of different approaches, but ended up with that of Fig. 1b, where the ϕ scalar field appears to be a good approximation of the spatial distribution of seismicity (Fig. 1a, b). Other choices for generating ϕ would lead to quantitatively different, but qualitatively similar results.

Extended Data Fig. 4 shows additional correlation, r, analysis based on smoothed seismicity. For this, we use an even-area regional sampling of ϕ and all scalar fields considered. As for the Molchan analysis, all fields are first smoothed with a $6\sigma = 100$ km spatial Gaussian filter, and statistical significance is estimated from 10,000 Monte Carlo simulations. We find a standard deviation for the random correlations of $\Sigma_r \approx 0.04$, and roughly normally distributed values. Extended Data Fig. 4 shows the $\pm 2\Sigma_r$ range outside which correlation would be inferred to be ~95% significant. We also conducted a shift test, displacing the rate of topography change field in a random direction by a distance of between 250 and 1,000 km and recomputing correlation with seismicity. From those tests, correlations of $|r| \geq 0.2$ and $|r| \geq 0.3$ are estimated to be 96% and 99.8% significant, respectively.

Extended Data Figure 4 indicates that Pearson correlation coefficient, r, results are very consistent with the relative model performance as indicated by the Molchan scores (Extended Data Fig. 2). We explored different smoothing; while absolute r values are affected, relative performance of models was found to be similar, as for S. For example, if we repeat the exercise of Extended Data Fig. 4 using a 250 km rather than 100 km Gaussian smoothing, $|r|$ is increased by ~0.05, but for all models by roughly the same amount.

Extended Data Figure 5 explores wavelength dependence of correlation further using sliding band-pass filtering, with details described in ref. 13. We find that the general wavelength-dependent behaviour is similar for both linear, r, and Spearman rank correlation, r_S, which provides confidence in the overall match between the signals. However, detailed variations, particularly for wavelengths $\lesssim 400$ km, are dependent on the type of correlation metric used. This indicates that total correlations in Extended Data Fig. 4 emphasize the differences between r and r_S by integration. It also implies that relatively small spatial scales are where neglected effects, such as the flexural rigidity of the lithosphere, and artefacts due to filtering, may predominantly complicate correlation analysis. However, the general wavelength dependence is irrespective of which type of correlation metric is used.

Extended Data Figure 5 indicates that the maximum shear strain rates from geodesy show the best correlation and form an upper envelope of wavelength-dependent match to seismicity. The geodynamically predicted rate change of dynamic topography rates also show strong, positive correlation, across most wavelengths $\gtrsim 200$ km. Moho depth gradients are mainly correlated with seismicity at ~500 km wavelengths, indicating that crustal complexities will impose some shorter wavelength modulations on the driving, mantle-based forcing. Finally, GPE gradients appear to contribute with positive correlation values throughout a range of wavelengths, but the total correlation values in Extended Data Fig. 4 are biased by the longest wavelength response, which may be misleading given the relatively small spatial extent of the study region.

Much of the intermountain seismicity band is clustered around Yellowstone itself (Fig. 1a, b), where we would expect shallow volcanic and magmatic processes to complicate any analysis. If we exclude a circular region of ~200 km diameter around Yellowstone, total correlations drop by $\Delta r \approx 0.1$ compared to Extended Data Fig. 4, but relative model performance is again confirmed.

Multivariable analysis. Correlation sampling also allows multivariable regression analysis, that is, solving a system of equations

$$A\mathbf{w} = \mathbf{\phi} \qquad (1)$$

for $\mathbf{w}$, the 'importance vector' with n dimensions for each of the subset of $n = 7$ models considered, in a least squares sense. $\mathbf{\phi}$ is the even-area sampled seismicity function with $m \approx 16{,}000$ entries where all n models are defined. (In practice, this covers ~90% of the region typically captured for Extended Data Fig. 4, leading to slightly different absolute r values.) A is m by n and holds the m predictions from the n models. Each column of A and $\mathbf{\phi}$ are demeaned and normalized to unity norm before solving equation (1) such that the elements of $\mathbf{w}$, w_i, are comparable in amplitude.

Extended Data Figure 7 shows the solution of equation (1); only the mantle-induced rate of dynamic topography change, GPE, and Moho and T_e gradients contribute to the non-negative, best-fit solution. The multivariable best fit yields $r \approx 0.52$ for regular and non-negative least squares, compared to the single, $r \approx 0.46$, solution. If we only use the four most important models, the least squares solution is naturally non-negative and also achieves $r \approx 0.52$.

An improvement in correlation of $\Delta r \approx 0.06$ for four compared to a single model is marginal, for example, compared to Σ_r. Part of the reason that no better fit to seismicity can be achieved is that the 'basis vectors' are, of course, not orthogonal, and individual models, such as Moho and T_e gradients, are themselves highly correlated. However, the interpretation of Extended Data Fig. 7 is consistent with a superposition of GPE and structural gradients on the rate change of topography from mantle flow as the main driver.

Comparison with vertical rates from GPS geodesy. We also evaluated the GPS velocities from the UNAVCO PBO solution (pbo.final_frame.vel at ftp://data-out.unavco.org/pub/products/velocity/). Extended Data Figure 1f shows a smoothed field constructed from the original vertical velocity estimates after rejection of outliers outside $\pm$ the RMS of the original data. Compared to the horizontals from GPS, vertical velocities are formally more uncertain, and, more importantly, strongly affected by hydrological, climatic, and anthropogenic signals. While the time series are improving, special processing is still required[45,46], and we consider distilling the tectonic and possible deep mantle signal from the GPS solutions to be work in progress.

However, initial results from a correlation analysis for regions away from the plate boundary indicate that rate change of dynamic topography from mantle flow shows poor total correlation with GPS verticals (Extended Data Fig. 8), if these are smoothed homogeneously, which may be inappropriate. Rate change of topography from mantle flow is positively correlated with GPS verticals within a band from ~250 km to ~800 km wavelength, however, similar to the peak of the Moho depth gradient correlation with seismicity in Extended Data Fig. 5. Only dynamic topography from flow, or GPE fields, show total, positive correlations with GPS verticals, and Moho depths are negatively correlated, meaning positive GPS verticals are found preferentially in regions of high GPE and thin crust.

If we compare the GPS verticals with horizontal kinematic constraints, we find that normal strain rates show positive correlation with GPS verticals, across a wide range of wavelengths (Extended Data Fig. 8). Such an association of change of dynamic topography with regions of extension is opposite to what might be expected for an edge-loaded, crustal thickening scenario, and rather more like extension on top of a region that is uplifted because of a basal density anomaly or push. This is consistent with a general expectation that, on scales of $\lesssim 200{\ldots}600$ km (ref. 15), strength of a visco-elasto-plastic lithosphere will partition any applied vertical normal stress primarily into membrane stress rather than uplift.

31. Levandowski, W., Jones, C. H., Shen, W., Ritzwoller, M. H. & Schulte-Pelkum, V. Origins of topography in the western U.S.: mapping crustal and upper mantle density variations using a uniform seismic velocity model. *J. Geophys. Res.* **119,** 2375–2396 (2014).
32. Liu, L. & Gurnis, M. Dynamic subsidence and uplift of the Colorado Plateau. *Geology* **38,** 663–666 (2010).
33. Moresi, L. N. & Solomatov, V. S. Numerical investigations of 2D convection with extremely large viscosity variations. *Phys. Fluids* **7,** 2154–2162 (1995).
34. Zhong, S., Zuber, M. T., Moresi, L. & Gurnis, M. Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *J. Geophys. Res.* **105,** 11063–11082 (2000).
35. Auer, L., Boschi, L., Becker, T. W., Nissen-Meyer, T. & Giardini, D. Savani: a variable-resolution whole-mantle model of anisotropic shear-velocity variations based on multiple datasets. *J. Geophys. Res.* **119,** 3006–3034 (2014).
36. Burov, E. & Guillou-Frottier, L. The plume head-continental lithosphere interaction using a tectonically realistic formulation for the lithosphere. *Geophys. J. Int.* **161,** 469–490 (2005).
37. Braun, J. The many surface expressions of mantle dynamics. *Nature Geosci.* **3,** 825–833 (2010).

38. Becker, T. W. On recent seismic tomography for the western United States. *Geochem. Geophys. Geosys.* **13,** Q01W10 (2012).
39. Schmandt, B. & Humphreys, E. Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth Planet. Sci. Lett.* **297,** 435–445 (2010).
40. Lou, X. & van der Lee, S. Observed and predicted North American teleseismic delay times. *Earth Planet. Sci. Lett.* **402,** 6–15 (2014).
41. Perry, H. K. C., Forte, A. M. & Eaton, D. W. S. Upper-mantle thermochemical structure below North America from seismic-geodynamic flow models. *Geophys. J. Int.* **154,** 279–299 (2003).
42. Moucha, R. *et al.* Deep mantle forces and the uplift of the Colorado Plateau. *Geophys. Res. Lett.* **36,** L19310 (2009).
43. NCEDC. *Northern California Earthquake Data Center: Dataset* (UC Berkeley Seismological Laboratory, 2014).
44. Engdahl, E. R., van der Hilst, R. D. & Buland, R. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull. Seismol. Soc. Am.* **88,** 722–743 (1998).
45. Serpelloni, E., Faccenna, C., Spada, G., Dong, D. & Williams, S. D. P. Vertical GPS ground motion rates in the Euro-Mediterranean region: new evidence of velocity gradients at different spatial scales along the Nubia-Eurasia plate boundary. *J. Geophys. Res.* **118,** 6003–6024 (2013).
46. Borsa, A. A., Agnew, D. C. & Cayan, D. R. Ongoing drought-induced uplift in the western United States. *Science* **345,** 1587–1590 (2014).
47. Shen, W., Ritzwoller, M. H. & Schulte-Pelkum, V. Crustal and uppermost mantle structure in the central U.S. encompassing the Midcontinent Rift. *J. Geophys. Res.* **118,** 4325–4344 (2013).
48. Obrebski, M., Allen, R. M., Pollitz, F. & Hung, S.-H. Lithosphere–asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities. *Geophys. J. Int.* **185,** 1003–1021 (2011).
49. James, D. E., Fouch, M. J., Carlson, R. W. & Roth, J. B. Slab fragmentation, edge flow and the origin of the Yellowstone hotspot track. *Earth Planet. Sci. Lett.* **311,** 124–135 (2011).
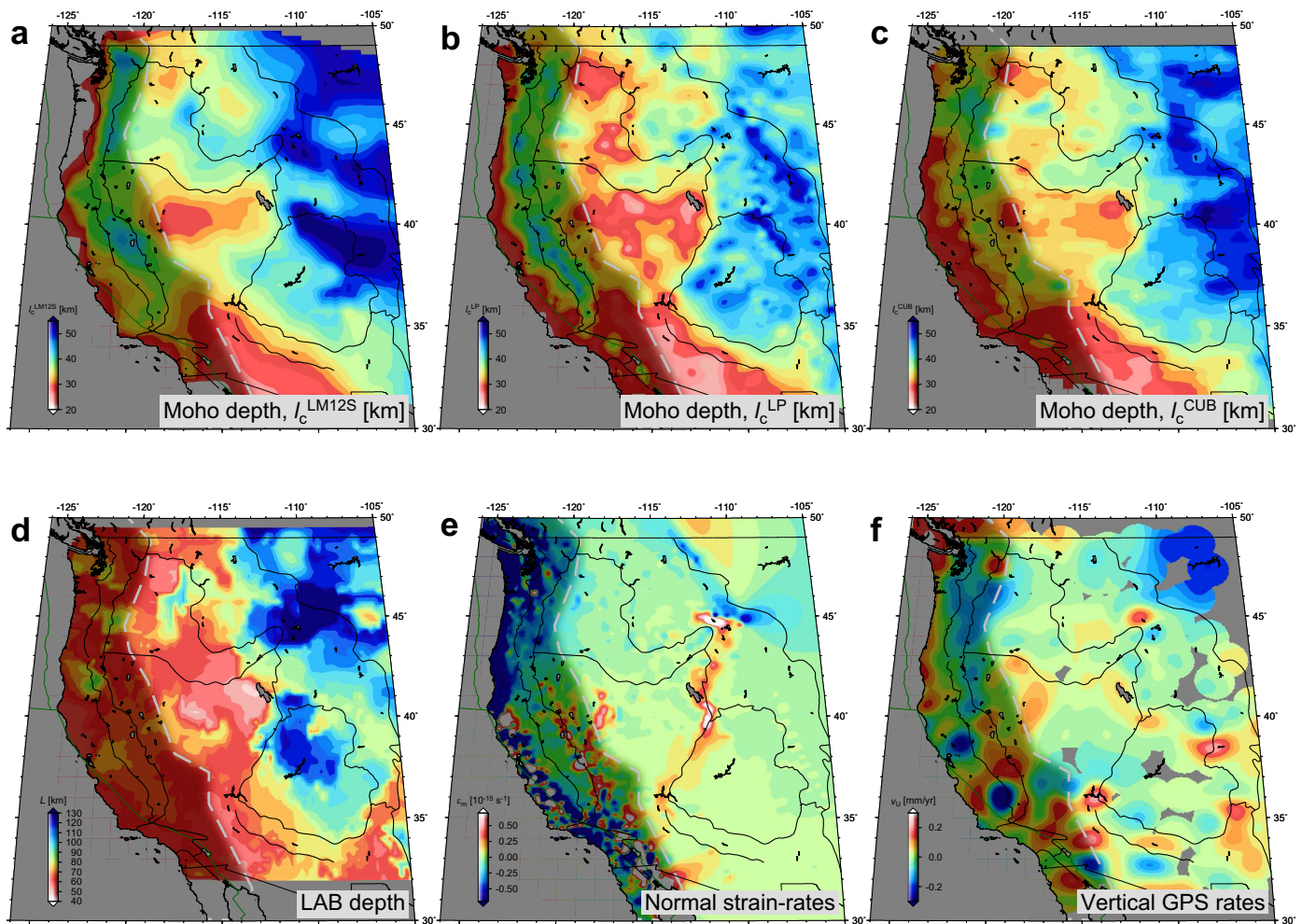
Extended Data Figure 1 | Maps of selected, additional scalar fields considered for the Molchan analysis in Fig. 2 and Extended Data Fig. 2. a–c, Moho depth estimates from Levander and Miller[3] (LM12S), Lowry and Pérez-Gussinyé[11] (LP), and Shen et al.[47] (CUB). **d–f**, 'Lithosphere asthenosphere boundary' (LAB) depth inferred from receiver functions by Levander and Miller[3], shear wave version (**d**), horizontal normal strainrates based on GPS geodesy from Kreemer et al.[29] (extension positive) (**e**), and vertical GPS rates, smoothed from the 2014 UNAVCO/PBO velocity release (pbo.final_frame.vel at ftp://data-out.unavco.org/pub/products/velocity/) with presumed outliers removed (**f**).
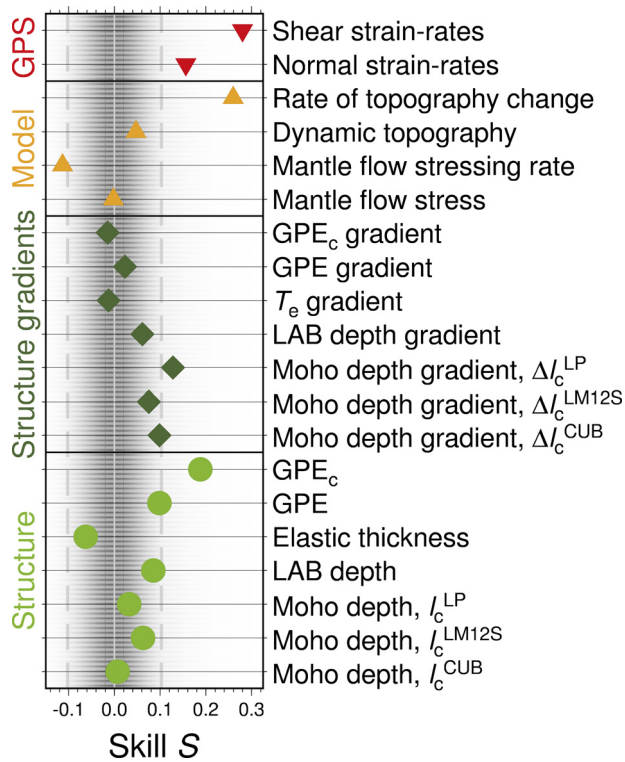
Extended Data Figure 2 | Molchan skill[20], S, for intraplate seismicity for an expanded set of models. We show results for three Moho models: CUB[47], LM12S[3] (S wave version) and LP[11] (as used in the main text, compare Extended Data Fig. 1a–c), estimates of LAB thickness from receiver functions (Extended Data Fig. 1d) and elastic thickness, and two kinds of GPE models, as in the main text, gradients thereof, geodynamic model predictions, and kinematic models for the western United States (Fig. 1). Dashed lines indicate the regions outside of which results are 95% significant. This more comprehensive analysis confirms the conclusions about relative model performance.
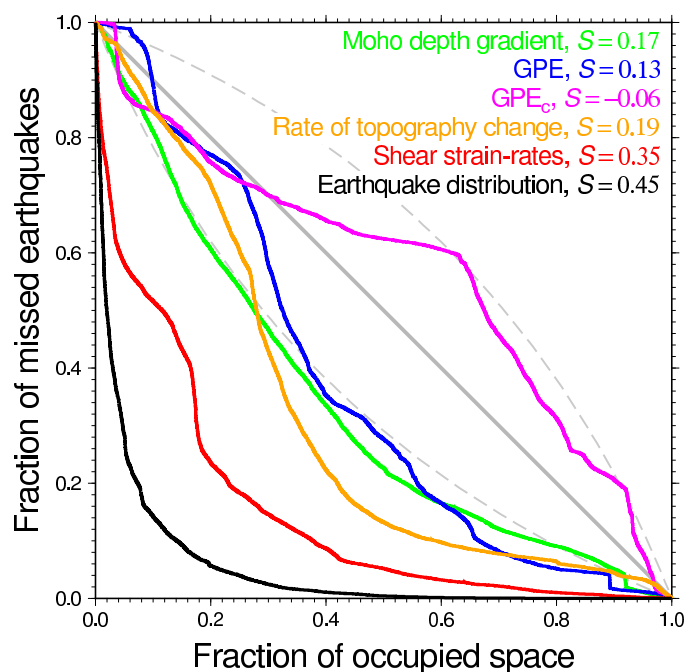
Extended Data Figure 3 | Molchan error curves for prediction of all seismicity in the entire region of Fig. 1b. Compare with Fig. 2a, which only considers seismicity outside the plate boundary zone as indicated in Fig. 1. Rate change of dynamic topography shows significant positive skill, even when including plate boundary seismicity.
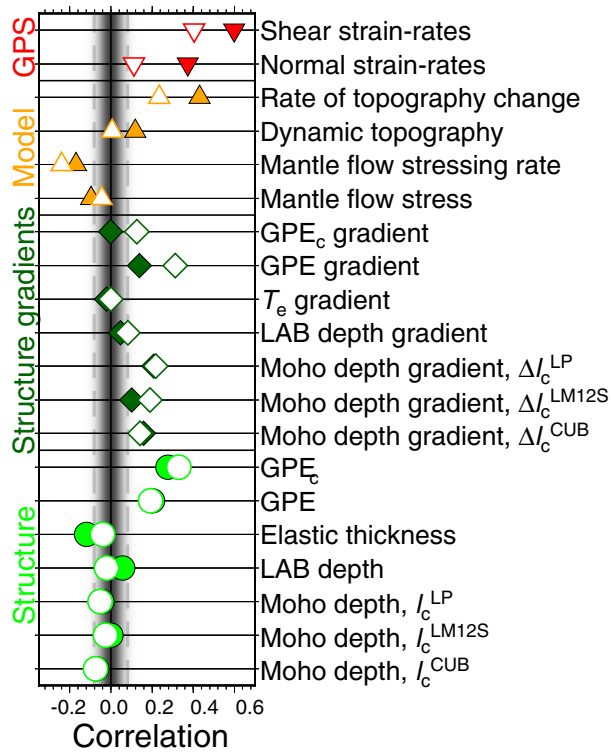
Extended Data Figure 4 | Correlation, *r*, between seismicity density (*ϕ*, Fig. 1b) and different models. Obtained by even-area sampling of regions outside the plate boundary zone in Fig. 1, compare with Extended Data Fig. 2. Filled and open symbols are for linear (Pearson), *r*, and Spearman rank, r_S, correlation, respectively. Dashed lines indicate the regions outside of which results are 95% significant, grey shading denotes approximate random density function distribution. Relative performance of Pearson correlation is consistent with results for skill (Extended Data Fig. 2), and Spearman rank shows some deviations, indicating lack of simple, linear relationships between *ϕ* and scalar fields.
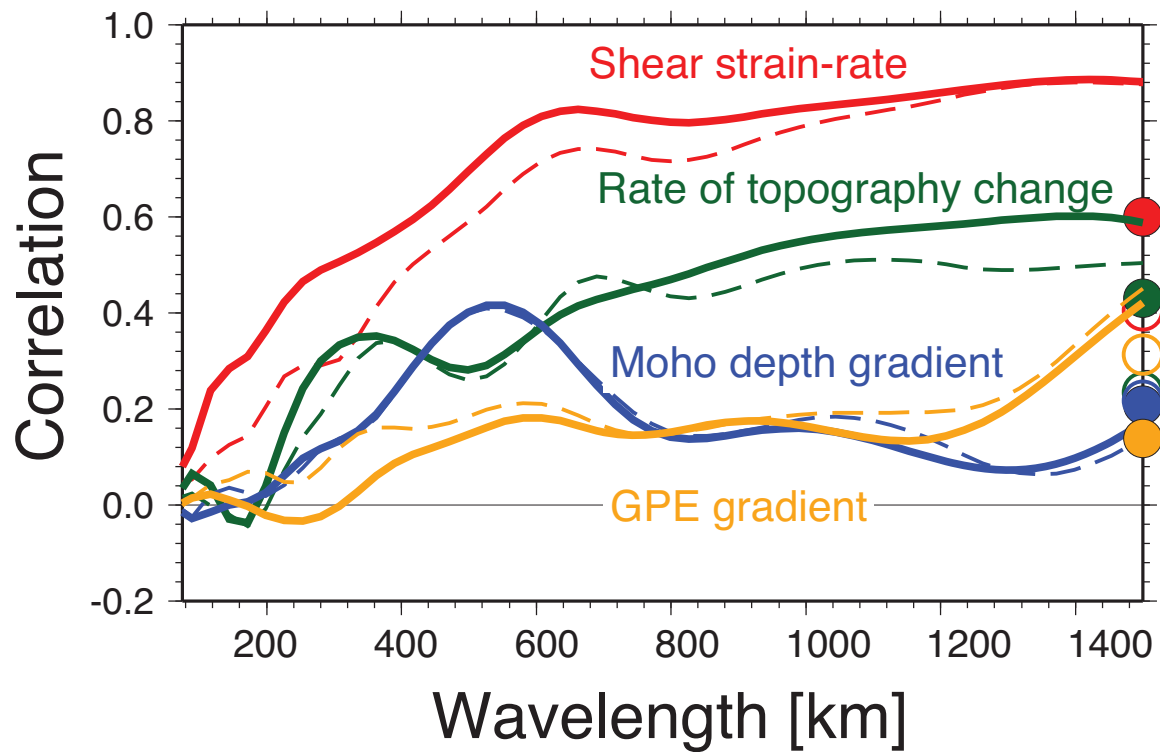
Extended Data Figure 5 | Wavelength-dependent correlation of seismicity and selected scalar fields. Models considered: shear strain rates from geodesy (Fig. 1c); rate change of dynamic topography (Fig. 1f); gradients of Moho depth[11]; and GPE (Fig. 1d). Circles on right show the total correlation as in Extended Data Fig. 4. Solid, heavy lines and filled symbols are for linear (Pearson) correlation, and thin, dashed lines and open symbols are for Spearman rank correlation, respectively. Rate change of dynamic topography shows positive correlation with seismicity across most wavelengths. Shorter wavelengths, where the response of a rheologically complex lithosphere will be most relevant, emphasize the differences between r and r_S seen in Extended Data Fig. 4.
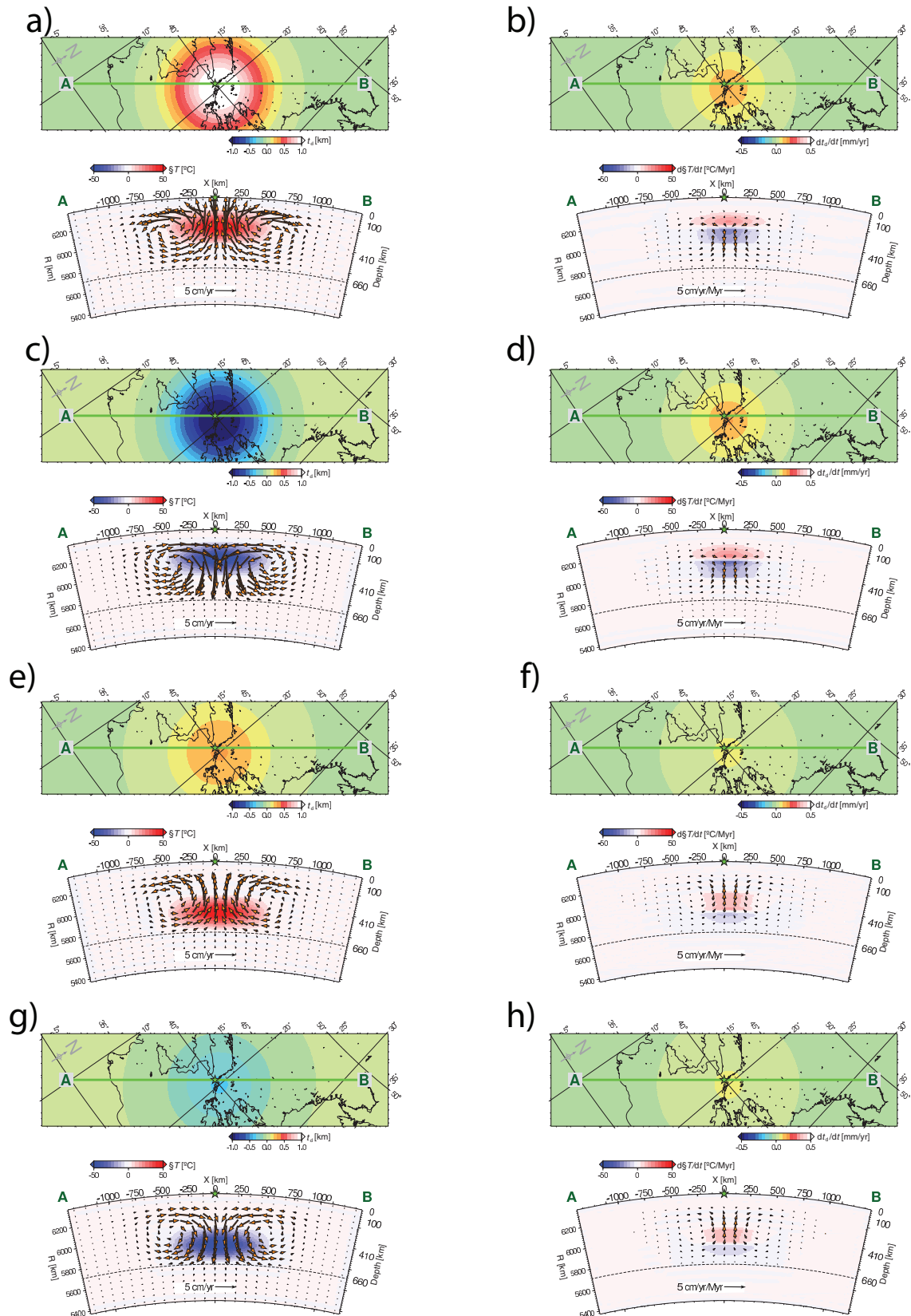
Extended Data Figure 6 | Illustrative, synthetic mantle flow models exploring dynamic topography and rate change thereof for simple, *ad hoc* temperature anomalies. a–h, Bottom subplots show temperature anomalies and mantle flow velocities, top subplots show dynamic topography on the surface (left) and temporal change thereof (right) for four mantle density cases along profiles. **a, b**, Shallow, hot anomaly; **c, d**, shallow, cold anomaly; **e, f**, deep, hot anomaly; **g, h**, deep, cold anomaly. Note how rate change of dynamic topography is positive for both cold, sinking and hot, rising anomalies because of reduction of negative and increase of positive dynamic topography, respectively[23,24].
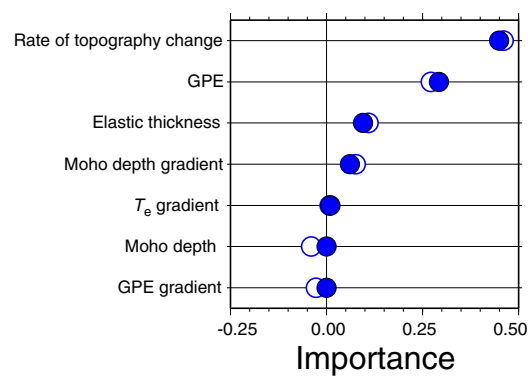
Extended Data Figure 7 | Multivariable regression analysis to best fit the seismicity density, evaluating a subset of the models of Extended Data Fig. 4. The normalized contribution of each model vector, the importance w_i, is indicated by filled and open symbols from a non-negative or regular least squares solution of equation (1), respectively. Results indicate that rate change of dynamic topography is the strongest driver of seismicity, with contributions from GPE, and modulated by Moho depth gradients and elastic thickness variations (compare with Extended Data Fig. 4).
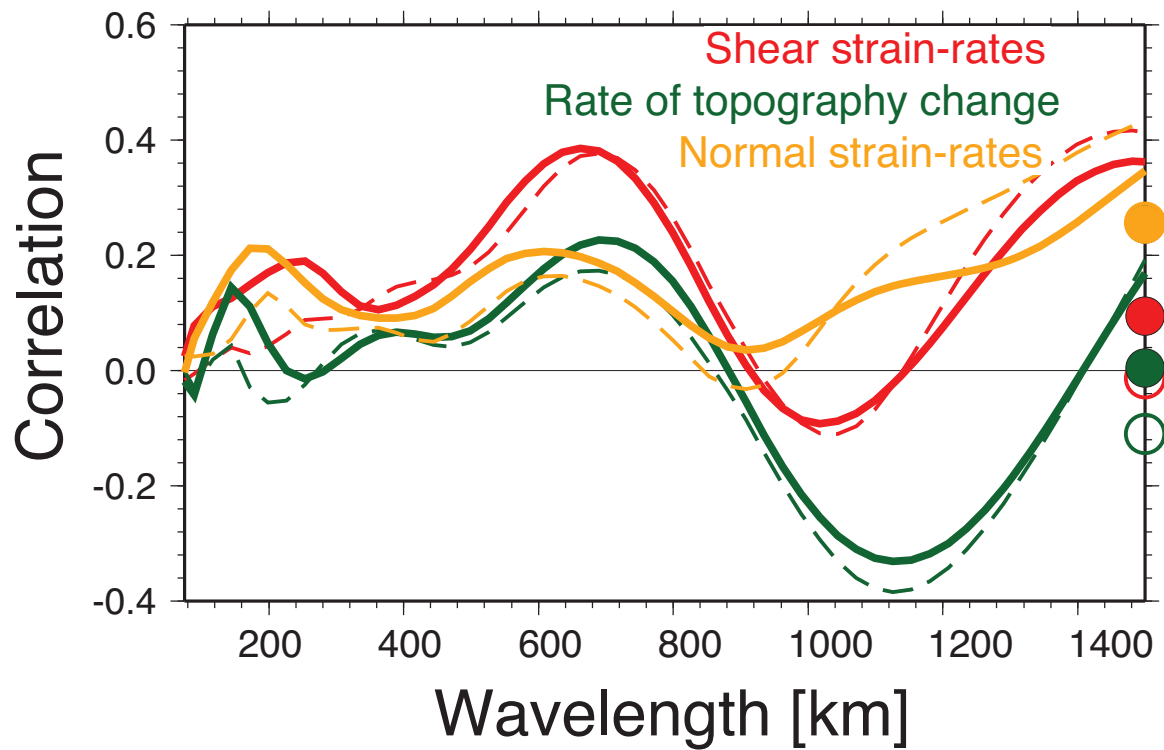
Extended Data Figure 8 | Wavelength-dependent correlation between vertical GPS rates (Extended Data Fig. 1f) and selected scalar fields. For models and legend see Extended Data Fig. 5. Rate of topography change and shear strain rates show positive correlation with geodetically imaged uplift at wavelengths of ~500...900 km, but rate change of dynamic topography is negatively correlated at longer wavelengths. Results indicate that rate change of dynamic topography may need to be understood more fully in context of a rheologically realistic model of lithospheric deformation, whereas geodetic uplift rates may be strongly affected by other signals, for example, climatic and hydrological.
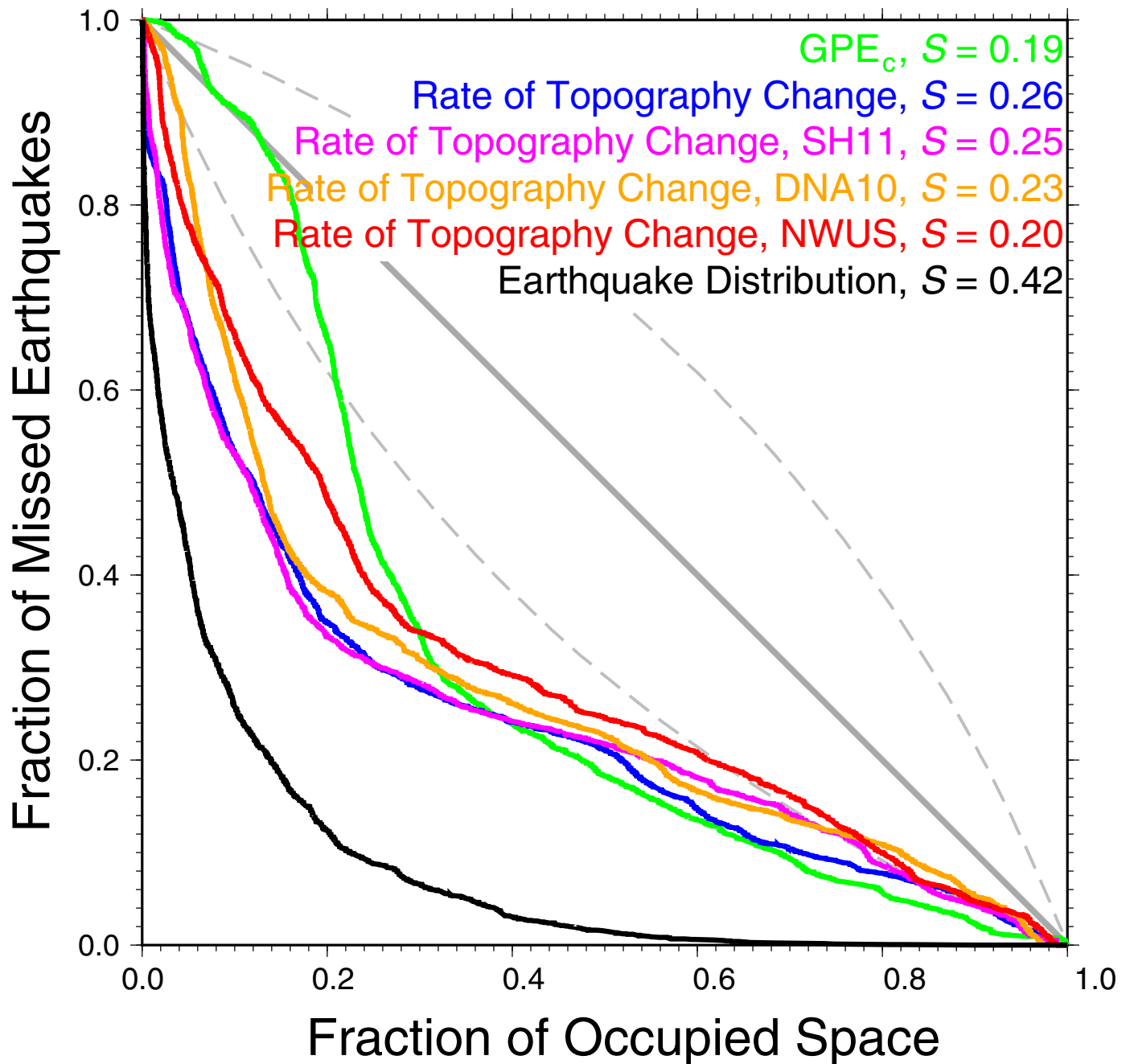
Extended Data Figure 9 | Molchan curves for mantle flow predictions for a range of tomographic models. Flow modelling uses our reference model[12] (compare with Fig. 2a and Extended Data Fig. 2), and older tomography[38]: SH11 (ref. 39) is a precursor of ref. 12; DNA10 a joint surface and body wave inversion by Obrebski *et al.*[48]; and NWUS-S an earlier, USArray-based shear wave tomography model by James *et al.*[49]. The best GPE model and smoothed seismicity are shown for reference. Older tomographic models substantiate our conclusions based on ref. 12; in particular SH11, which was found to bestpredict delay times[40], has near-identical skill.

LETTER

The pre–vertebrate origins of neurogenic placodes

Philip Barron Abitua[1]†, T. Blair Gainous[1]†, Angela N. Kaczmarczyk[1], Christopher J. Winchell[1], Clare Hudson[2], Kaori Kamata[3], Masashi Nakagawa[3], Motoyuki Tsuda[3], Takehiro G. Kusakabe[4] & Michael Levine[1]†

The sudden appearance of the neural crest and neurogenic placodes in early branching vertebrates has puzzled biologists for over a century[1]. These embryonic tissues contribute to the development of the cranium and associated sensory organs, which were crucial for the evolution of the vertebrate "new head"[2,3]. A previous study suggests that rudimentary neural crest cells existed in ancestral chordates[4]. However, the evolutionary origins of neurogenic placodes have remained obscure owing to a paucity of embryonic data from tunicates, the closest living relatives to those early vertebrates[5]. Here we show that the tunicate *Ciona intestinalis* exhibits a proto-placodal ectoderm (PPE) that requires inhibition of bone morphogenetic protein (BMP) and expresses the key regulatory determinant *Six1/2* and its co-factor *Eya*, a developmental process conserved across vertebrates. The *Ciona* PPE is shown to produce ciliated neurons that express genes for gonadotropin-releasing hormone (*GnRH*), a G-protein-coupled receptor for relaxin-3 (*RXFP3*) and a functional cyclic nucleotide-gated channel (*CNGA*), which suggests dual chemosensory and neurosecretory activities. These observations provide evidence that *Ciona* has a neurogenic proto-placode, which forms neurons that appear to be related to those derived from the olfactory placode and hypothalamic neurons of vertebrates. We discuss the possibility that the PPE-derived GnRH neurons of *Ciona* resemble an ancestral cell type, a progenitor to the complex neuronal circuit that integrates sensory information and neuroendocrine functions in vertebrates.

Neurogenic placodes contribute to cranial sensory systems that mediate hearing, smell and taste[2]. They are considered a vertebrate innovation[6] and the full repertoire, including olfactory, otic, epibranchial and trigeminal placodes, are already present in jawless hagfish and lampreys. It has been argued that the evolution of neurogenic placodes was a crucial event for facilitating the transition from filter-feeding invertebrate chordates to predatory vertebrates[3].

Comparative embryological studies have failed to identify any clear evolutionary precursors of neurogenic placodes in invertebrate chordates. It has been suggested that the pre-oral organ of amphioxus, Hatschek's pit, is related to the adenohypophyseal (anterior pituitary) placode of vertebrates; however, the fate map of this tissue remains unresolved[7]. In addition, the atrial primordia of *Ciona* have been likened to the otic and lateral-line placodes of fish, but they are not specified until tailbud stages and do not derive from the anterior neural plate border[8]. Also, previous studies provided evidence for a placodal-like tissue in *Ciona*, but it was homologized to the adenohypophyseal placode[8,9], which lacks neurogenic potential. We show that the latter tissue does in fact produce ciliated sensory neurons that express *GnRH* and *RXFP3*, signatures of neuroendocrine cell types found in vertebrates[10–12].

Neurogenic placodes develop from a specific domain anterior to the neural plate of vertebrate embryos[2]. All placode subtypes arise from this U-shaped domain, termed the preplacodal ectoderm[13]. This tissue expresses the transcription factor *Six1*, and mutations in *Six1* cause

craniofacial defects in humans, including deafness[14]. We found that the *Ciona* homologue of *Six1*, *Six1/2*, as well as its co-factor *Eya*, is initially expressed in a row of eight cells located immediately anterior to the neural plate, as marked by *ZicL* (Fig. 1a and Extended Data Fig. 1). This pattern is recapitulated by a *Six1/2* regulatory sequence driving expression of an *mCherry* reporter gene (*Six1/2>mCherry*;
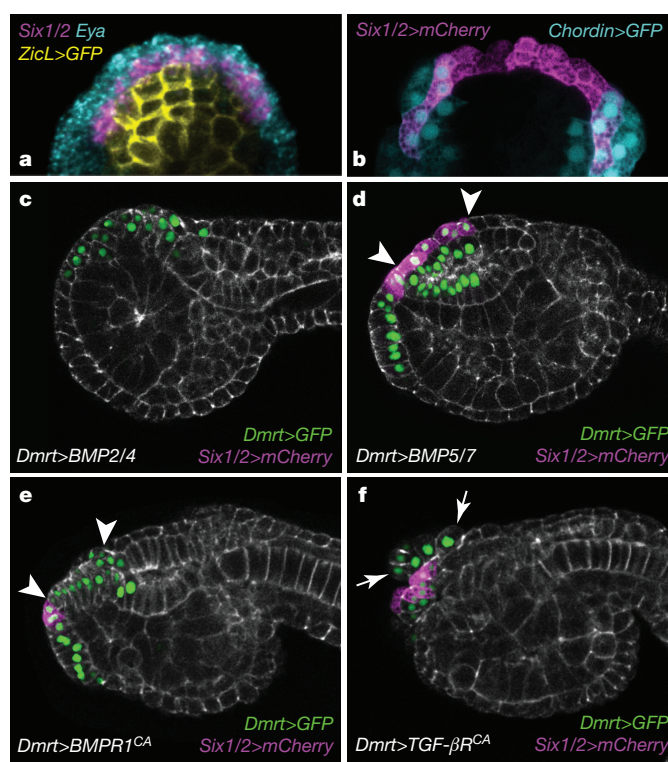

Figure 1 | *Six1/2* expression requires BMP attenuation. a, Dorsal view of an embryo at early neurula stage electroporated with *ZicL>GFP* and hybridized with *Six1/2* and *Eya* mRNA probes. **b**, Dorsal view of an embryo at mid-neurula stage co-electroporated with *Six1/2>mCherry* and *Chordin>GFP* after an additional division (16 Six1/2⁺ cells in total). **c–f**, Tailbud electroporated with *Six1/2>mCherry* and *Dmrt>GFP*; counterstained with phalloidin (grey). **c**, Tailbud co-electroporated with *Dmrt>BMP2/4* (131 of 135 embryos showed no *Six1/2>mCherry* expression). **d**, Tailbud co-electroporated with *Dmrt>BMP5/7* (120 of 135 embryos displayed a full expression pattern for *Six1/2>mCherry*). **e**, Tailbud co-electroporated with *Dmrt>BMPR1^{CA}* (95 of 135 embryos had partial anterior *Six1/2>mCherry* expression). **f**, Tailbud co-electroporated with *Dmrt>TGF-βR^{CA}* (93 of 135 embryos displayed a full expression pattern for *Six1/2>mCherry* and anterior neural tube defects). Arrowheads in **d** and **e** flank the derivatives of the anterior ectoderm that normally co-express *Six1/2>mCherry* and *Chordin>GFP* in control embryos (see panel **b**). Arrows show extrusion of sensory vesicle cells in **f**.

[1]Center for Integrative Genomics, Division of Genetics, Genomics and Development, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [2]Sorbonne Universités, Université Pierre et Marie Curie, Centre National de la Recherche Scientifique, Laboratoire de Biologie du Développement de Villefranche-sur-mer, Observatoire Océanologique, 06230 Villefranche-sur-mer, France. [3]Graduate School of Life Science, University of Hyogo, Kamigori, Hyogo 678-1297, Japan. [4]Institute for Integrative Neurobiology and Department of Biology, Faculty of Science and Engineering, Konan University, Kobe 658-8501, Japan. †Present addresses: Department of Molecular and Cellular Biology, Harvard University, Massachusetts 02138, USA (P.B.A.); Cardiovascular Research Institute, University of California, San Francisco, California 94158, USA (T.B.G.); Lewis-Sigler Institute of Integrative Genomics, Princeton University, New Jersey 08544, USA (M.L.).

Fig. 1b and Supplementary Video 1). At tailbud stages the Six1/2[+] cells intercalate and become situated above the anterior brain (that is, the sensory vesicle), surrounding the future oral opening (Extended Data Fig. 2).

In vertebrates, the preplacodal ectoderm is specified by a two-step process during which a gradient of BMP signalling establishes preplacodal competence in the non-neural ectoderm during the blastula/early gastrula stage, followed by subsequent inhibition of BMP signalling during gastrulation[13]. BMP signalling may not be required for establishing proto-placodal competence in *Ciona*, but inhibition of BMP signalling seems to be crucial for its specification. We observe conserved expression of *BMP2/4* and the *Chordin* (also known as *Ci-Chordin*) antagonist in ventral and dorsal regions, respectively, at the time when the PPE is specified during gastrulation and neurulation (Extended Data Fig. 3).

Localized *Chordin* expression at the lateral edges of the *Ciona* PPE (Fig. 1b) suggests a role in antagonizing BMPs emanating from ventral regions, similar to the situation seen in vertebrates[15]. We therefore explored the possibility that inhibition of BMP signalling might be required for the activation or maintenance of *Six1/2* expression. We sought to overcome the presumed inhibitory effect of chordin by the targeted misexpression of BMP ligands or constitutively active forms of BMP receptors. These experiments were performed using a previously described *Dmrt* regulatory DNA (that is, *Dmrt>GFP*)[16], which drives expression throughout the anterior neural plate and adjacent anterior ectoderm (Extended Data Fig. 1).

The co-electroporation of *Dmrt>BMP2/4* with *Dmrt>GFP* and *Six1/2>mCherry* resulted in loss of *Six1/2* expression in the anterior neural plate border, consistent with similar experiments in vertebrates[17] (Fig. 1c). Conversely, misexpression of *BMP5/7* had no noticeable effect on expression of the *Six1/2* reporter or embryo morphology (Fig. 1d). These observations suggest that chordin may be specifically required to exclude BMP2/4 signals from the PPE.

Additional misexpression assays were performed with a constitutively active form of the BMP receptor (BMPR1[CA]) for cell-autonomous activation of BMP signalling. In these experiments, activation is restricted to Dmrt[+] cells, which form the anterior neural plate and adjacent anterior ectoderm. Ectopic BMPR1 activity results in reduced expression of *Six1/2* (Fig. 1e), similar to the results obtained upon misexpression of the BMP2/4 ligand. However, the phenotype is less severe, with persistence of *Six1/2* expression in the medially derived regions of the PPE (compare Fig. 1d with Fig. 1e). Expression is lost in lateral regions, which are the source of the GnRH neurons.

To demonstrate that disrupting PPE formation is a specific result of ectopic BMP signalling, we misexpressed a constitutively active form of the transforming growth factor-β receptor (TGF-βR[CA]) throughout the anterior brain and adjacent ectoderm (*Dmrt>TGF-βR[CA]*). This had very little effect on the expression of *Six1/2*, despite severe defects in brain morphology (Fig. 1f). We therefore conclude that inhibition of BMP signalling is essential for the specification of the *Ciona* PPE, as observed in vertebrates[15,17]. However, it is possible that this inhibition indirectly regulates *Six1/2* expression.

The olfactory placode is the anterior-most neurogenic placode in vertebrates and develops above the oral opening[7]. GnRH neurons arising from this placode migrate along the axonal tracts of olfactory neurons towards the hypothalamus[11,12]. They subsequently innervate the anterior pituitary and stimulate the release of gonadotropins, hormones critical for growth, sexual development and reproduction[18]. Previous studies have documented the expression of six different GnRH peptides encoded by two duplicated genes in *Ciona*[19]. *GnRH2* (hereafter referred to as *GnRH*) is expressed near the oral opening (Extended Data Fig. 4) and we used the 5′ regulatory sequences (*GnRH>GFP*) to label this lineage during development. During the tailbud stage, the posterior-most Six1/2[+] cells acquire a distinctive cone-shaped morphology, permitting their direct visualization during morphogenesis (Fig. 2a and Extended Data Fig. 2). By the larval
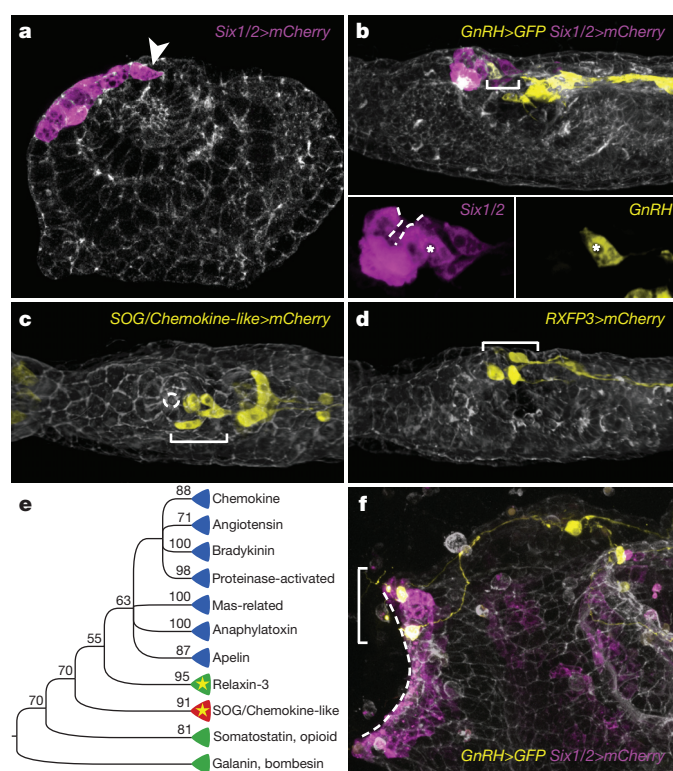

Figure 2 | PPE-derived GnRH neurons express *RXFP3*. a, Tailbud electroporated with *Six1/2>mCherry*. Arrowhead indicates the posterior-most Six1/2[+] cell that undergoes neurogenesis. **b**, Larva co-electroporated with *GnRH>GFP* and *Six1/2>mCherry*. Insets show *Six1/2* expression (left) and *GnRH* expression (right) surrounding the oral opening (dotted line). The GnRH[+] neuron is marked with an asterisk. **c**, Dorsal view of larva electroporated with *SOG/Chemokine-like>mCherry*. The dotted line indicates the oral opening. **d**, Larva electroporated with *RXFP3>mCherry*. **e**, Phylogenetic placement among human rhodopsin-class GPCRs of *Ciona* *RXFP3* and *SOG/Chemokine-like*, indicated by the starred clades. Green clades comprise *Ciona* and human orthologues, whereas no human orthologue was found for the red clade and no known *Ciona* GPCR grouped within the blue clades. Support values are Bayesian posterior probability percentages. **f**, Juvenile co-electroporated with *GnRH>GFP* and *Six1/2>mCherry*. The dotted line indicates the oral opening. *Ciona* were counterstained with phalloidin (grey) in panels **a**–**d** and **f**. Brackets in panels **b**–**d** and **f** indicate PPE-derived GnRH neurons. Panels **a**, **b**, **d** and **f** are lateral anterior views.

stage, these cells have differentiated into neurons that co-express *GnRH>GFP* and *Six1/2>mCherry* (Fig. 2b and Supplementary Video 2). The production of these neurons is strongly inhibited by BMP signalling (Extended Data Fig. 5).

Ciona PPE-derived GnRH neurons express two genes encoding G-protein-coupled receptors (GPCRs) implicated in neurosecretion in vertebrates, a somatostatin/opioid/galanin/chemokine-like receptor (*SOG/Chemokine-like*) and a relaxin-3 receptor (*RXFP3*) (Fig. 2c, d, Extended Data Fig. 4 and Supplementary Video 2). The relationship of these *Ciona* GPCRs to human rhodopsin-class GPCRs is shown in a summary tree (Fig. 2e and Extended Data Figs 6 and 7). In vertebrates, *RXFP3* is expressed in hypothalamic paraventricular nuclei that provide synaptic inputs for the secretion of GnRH within the hypothalamic–pituitary–gonadal (HPG) axis[20]. Notably, the PPE-derived *Ciona* GnRH neurons survive through metamorphosis despite extensive cell death and reorganization of most of the larval nervous system (Fig. 2f). Moreover, injection of GnRH into mature *Ciona* causes the release of gametes and is likely to trigger spawning in wild populations[21].

As both GnRH1 and chemosensory neurons arise from the olfactory placode of vertebrates, we were curious whether *Ciona* PPE-derived neurons might also possess chemosensory properties. Chemosensation by olfactory neurons occurs via odorant receptors in the tips of their
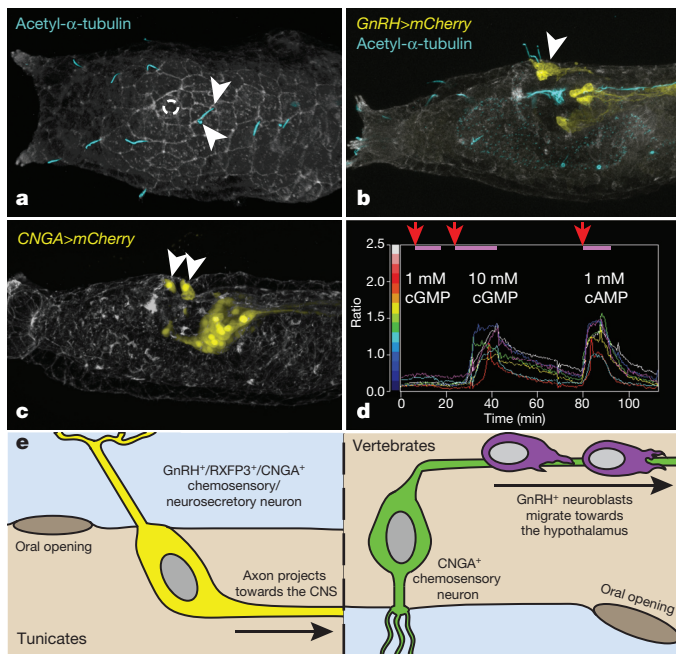
Figure 3 | Ciliated GnRH neurons express functional *CNGA*. a–c, Larvae stained with phalloidin (grey). **a**, Dorsal view of a larva stained with an acetylated α-tubulin antibody; dotted line indicates the oral opening. **b**, Lateral view of a larva electroporated with *GnRH>mCherry* and stained with an acetylated α-tubulin antibody. **c**, Lateral view of a larva electroporated with *CNGA>mCherry*. Arrowheads indicate PPE-derived GnRH neurons in all panels. **d**, Ratiometric calcium imaging of HEK 293 cells transfected with *Ciona* CNGA, assayed by the intracellular Ca^{2+} indicator Fura-2. The graph shows the efficiency of response of CNGA to the cyclic nucleotide analogues 8-Br-cGMP and 8-Br-cAMP in seven individual cells (coloured lines) over a 120-min period. Indicated are the time points (red arrows) and durations (pink bars) of each analogue perfusion. **e**, Schematic that compares the PPE-derived GnRH neurons in tunicates such as *Ciona* (left) and olfactory-derived neurons in vertebrates (right). CNS, central nervous system; HEK, human embryonic kidney.

cilia. Swimming larvae possess a group of neurons located behind the oral opening, previously described as the anterior apical trunk epidermal neurons (aATENs). These neurons possess cilia that are clearly visualized using an acetylated α-tubulin antibody (Fig. 3a). Electroporation with *GnRH>mCherry* followed by acetylated α-tubulin staining confirmed that the cilia project from the PPE-derived GnRH neurons (Fig. 3b). Moreover, comprehensive visualization of Six1/2$^+$ cell morphogenesis reveals that the aATEN–GnRH neurons are derived from the a11.205 lineage (Extended Data Fig. 2). Ultrastructural analysis has shown that cilia of these neurons have a 9+2 microtubule arrangement[22], a shared characteristic attributed to non-motile cilia of vertebrate olfactory neurons[23].

Conventional odorant receptors have not been identified in the *Ciona* genome[7], but it is unlikely that a marine chordate would lack a means to detect odorants. Instead of examining divergent GPCRs, we focused on downstream effectors of chemosensory transduction. Cyclic nucleotide-gated (CNG) cation channels are critical for mediating sensory perception in photoreceptors and olfactory neurons. The binding of odorants to receptors causes an intracellular increase in cyclic nucleotides that, in turn, results in an influx of calcium, which ultimately leads to the depolarization of sensory neurons[24]. In rats, *CNGA2* expression is detected in GnRH1 neurons but is not necessary for their activity. This expression might be due to prior activation in the shared progenitor of GnRH1 and olfactory neurons[25,26].

A homologue of vertebrate *CNGA1* is expressed in fully differentiated aATEN–GnRH neurons of *Ciona* (Fig. 3c). To determine whether *Ciona* CNGA forms a functional channel, we transfected a

CNGA expression construct into human embryonic kidney 293 cells incubated with the intracellular calcium indicator Fura-2 AM to test for a response to cyclic nucleotides. Indeed, application of either 8-bromoguanine-3′,5′-cyclic monophosphate (8-Br-cGMP) or 8-bromoadenosine-3′,5′-cyclic monophosphate (8-Br-cAMP) elicited an increase in intracellular calcium concentration, with a higher sensitivity to cAMP than cGMP (Fig. 3d and Extended Data Fig. 8). The occurrence of a functional CNGA in PPE-derived neurons, in addition to the presence of a 9+2 cilium, suggests that they are likely to possess chemosensory activities. These observations provide the first evidence, to our knowledge, that putative sensory neurons arise from a placodal-like territory in a non-vertebrate.

Our data suggest that the aATEN–GnRH neurons have dual neurosecretory and chemosensory properties. In vertebrates, neurosecretory and chemosensory cells both arise from the olfactory placode and are intimately linked. GnRH neuroblasts use the axon tracts of chemosensory neurons to guide them to their final destination in the hypothalamus. In adults, pheromones detected by chemosensory neurons cause a release of gonadotropins via hypothalamic GnRH neurons, a function crucial to sexual reproduction. Furthermore, transneuronal tracing experiments have shown that olfactory and GnRH neurons form a coherent neuronal circuit[27,28], a connection that likely arose long ago.

It has been proposed that neuronal circuits evolve from the functional segregation of specialized cell types from multifunctional ancestral cells, akin to the well-known mechanism of gene duplication and subfunctionalization. A chemosensory–neurosecretory precursor is likely to have existed in ancestral chordates before the diversification of the HPG axis seen in vertebrates[29,30]. We have presented evidence that the PPE-derived aATEN–GnRH neurons in *Ciona* provide a particularly vivid example of such an ancestral cell type. Over the course of evolution within the vertebrate lineage, the proposed chemosensory and neurosecretory functions of these neurons may have become segregated into dedicated cell types that work together within a coherent circuit (Fig. 3e). Such cellular subfunctionalization might be a generally important mechanism of neuronal circuit evolution in vertebrates.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Gaskell, W. H. *The Origin of Vertebrates* (Longmans, Green & Co., 1908).
2. Steventon, B., Mayor, R. & Streit, A. Neural crest and placode interaction during the development of the cranial sensory system. *Dev. Biol.* **389,** 28–38 (2014).
3. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: a new head. *Science* **220,** 268–273 (1983).
4. Abitua, P. B., Wagner, E., Navarrete, I. A. & Levine, M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature* **492,** 104–107 (2012).
5. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439,** 965–968 (2006).
6. Shimeld, S. M. & Holland, P. W. Vertebrate innovations. *Proc. Natl Acad. Sci. USA* **97,** 4449–4452 (2000).
7. Patthey, C., Schlosser, G. & Shimeld, S. M. The evolutionary history of vertebrate cranial placodes – I: cell type evolution. *Dev. Biol.* **389,** 82–97 (2014).
8. Mazet, F. *et al.* Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. *Dev. Biol.* **282,** 494–508 (2005).
9. Christiaen, L., Bourrat, F. & Joly, J. A modular *cis*-regulatory system controls isoform-specific *pitx* expression in ascidian stomodæum. *Dev. Biol.* **277,** 557–566 (2005).
10. Sutton, S. W. *et al.* Distribution of G-protein-coupled receptor (GPCR) 135 binding sites and receptor mRNA in the rat brain suggests a role for relaxin-3 in neuroendocrine and sensory processing. *Neuroendocrinology* **80,** 298–307 (2004).
11. Schwanzel-Fukuda, M. & Pfaff, D. W. Origin of luteinizing hormone-releasing hormone neurons. *Nature* **338,** 161–164 (1989).
12. Wray, S., Grant, P. & Gainer, H. Evidence that cells expressing luteinizing hormone-releasing hormone mRNA in the mouse are derived from progenitor cells in the olfactory placode. *Proc. Natl Acad. Sci. USA* **86,** 8132–8136 (1989).
13. Kwon, H.-J., Bhat, N., Sweet, E. M., Cornell, R. A. & Riley, B. B. Identification of early requirements for preplacodal ectoderm and sensory organ development. *PLoS Genet.* **6,** e1001133 (2010).

14. Ruf, R. G. *et al.* SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1–SIX1-DNA complexes. *Proc. Natl Acad. Sci. USA* **101,** 8090–8095 (2004).
15. Reichert, S., Randall, R. A. & Hill, C. S. A BMP regulatory network controls ectodermal cell fate decisions at the neural plate border. *Development* **140,** 4435–4444 (2013).
16. Wagner, E. & Levine, M. FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139,** 2351–2359 (2012).
17. Ahrens, K. & Schlosser, G. Tissues and signals involved in the induction of placodal *Six1* expression in *Xenopus laevis. Dev. Biol.* **288,** 40–59 (2005).
18. Wray, S. From nose to brain: development of gonadotrophin-releasing hormone-1 neurones. *J. Neuroendocrinol.* **22,** 743–753 (2010).
19. Kusakabe, T. G. *et al.* A conserved non-reproductive GnRH system in chordates. *PLoS ONE* **7,** e41955 (2012).
20. McGowan, B. M. *et al.* Relaxin-3 stimulates the hypothalamic-pituitary-gonadal axis. *Am. J. Physiol. Endocrinol. Metab.* **295,** E278–E286 (2008).
21. Terakado, K. Induction of gamete release by gonadotropin-releasing hormone in a protochordate, *Ciona intestinalis. Gen. Comp. Endocrinol.* **124,** 277–284 (2001).
22. Konno, A. *et al.* Distribution and structural diversity of cilia in tadpole larvae of the ascidian *Ciona intestinalis. Dev. Biol.* **337,** 42–62 (2010).
23. Reese, T. S. Olfactory cilia in the frog. *J. Cell Biol.* **25,** 209–230 (1965).
24. Kaupp, U. B. Olfactory signalling in vertebrates and insects: differences and commonalities. *Nature Rev. Neurosci.* **11,** 188–200 (2010).
25. Constantin, S. & Wray, S. Gonadotropin-releasing hormone-1 neuronal activity is independent of cyclic nucleotide-gated channels. *Endocrinology* **149,** 279–290 (2008).
26. El-Majdoubi, M. & Weiner, R. I. Localization of olfactory cyclic nucleotide-gated channels in rat gonadotropin-releasing hormone neurons. *Endocrinology* **143,** 2441–2444 (2002).
27. Boehm, U., Zou, Z. & Buck, L. B. Feedback loops link odor and pheromone signaling with reproduction. *Cell* **123,** 683–695 (2005).
28. Yoon, H., Enquist, L. W. & Dulac, C. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123,** 669–682 (2005).
29. Gorbman, A. Olfactory origins and evolution of the brain-pituitary endocrine system: facts and speculation. *Gen. Comp. Endocrinol.* **97,** 171–178 (1995).
30. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nature Rev. Genet.* **9,** 868–882 (2008).

METHODS

Embryo preparation and imaging. *Ciona intestinalis* adults were obtained, fertilized *in vitro* and electroporated for transient transgenesis as previously described[16]. For each electroporation, typically 70 μg of DNA was resuspended in 100 μl buffer. Embryos were fixed at the appropriate developmental stage for 15 min in 4% formaldehyde. The tissue was then cleared in a series of washes of 0.01% Triton-X in PBS. Actin was stained overnight with Alexa-647-conjugated phalloidin (Fisher, A22287) at a dilution of 1:500.

For the acetyl-α-tubulin staining, larvae were fixed for 10 min in 2% paraformaldehyde, then washed three times in 0.01% Triton-X in PBS for 10 min each, and blocked with 1% normal goat serum for 10 min. Afterwards samples were incubated with a monoclonal anti-acetylated tubulin antibody (Sigma-Aldrich, T6793) at a dilution of 1:1,000 overnight at 4 °C. Next, the tissue was washed in 0.01% Triton-X in PBS, three times for 30 min each. The samples were then incubated for 2 h at room temperature with an Alexa Fluor 488 donkey anti-mouse IgG (H+L) antibody (Fisher, A21202) at a dilution of 1:500 in 1% normal goat serum, and finally washed in 0.01% Triton-X in PBS (three times for 30 min each).

Samples were mounted in 50% glycerol in PBS, with 2% 1,4-diazabicyclo[2,2,2]octane (Sigma) for microscopy. Confocal images were acquired on a Zeiss LSM 700 microscope using a Plan-Apochromat 20×, 40× or 63× objective. Confocal stacks contained approximately 50 optical slices at a thickness of 1–2 μm each. Images were rendered using Volocity 6 with the 3D opacity, extended focus or *XY* plane visualization tools. Time-lapse images were taken on a Zeiss LSM 700 microscope at intervals of 3–4 min. Bright field *in situ* micrographs of larvae were acquired on a Zeiss Axio Imager microscope using an EC Plan-Neofluar 40× objective with a ProgRes C14 plus camera.

Experimental perturbations were performed in triplicate and two biological replicates were scored. No statistical methods were used to predetermine sample size. The scored replicates were totalled and these are indicated in the appropriate figure legends. Phenotypes were scored in animals exhibiting transgenesis (that is, using a reporter unaffected by a given perturbation such as *Dmrt>GFP* in Fig. 1). Unfertilized animals or those that failed to gastrulate were excluded from the analysis. No randomization or blinding was performed on samples.

Molecular cloning. The Genome Browser Gateway (University of California, Santa Cruz) facilitated the identification of conserved non-coding sequences between *Ciona intestinalis* and *Ciona savignyi*. Putative enhancer sequences were amplified by polymerase chain reaction (PCR) and cloned into a pCESA vector using AscI/NotI restriction sites for *Chordin* (KH.C6.145; 5′-CGGTTACGTT AAGTTCGTGCG-3′ and 5′-CTTTGTTCCTGCTCGAAGTGAG-3′), *Six1/2* (KH.C3.553; 5′-GTGACAGGAAACGTCTAGG-3′ and 5′-TGGCTCTGAGG CCGTGATTGTAG-3′), *GnRH* (KH.C9.484; 5′-GTGTACGATTGACAAGT TGG-3′ and 5′-TGTTACGTTATCTCTCTAGAAG-3′), *SOG/Chemokine-like* (ci0100133186; 5′-GGCAGATTTGACCTCAACTTG-3′ and 5′-GGCGTTTCC GAAAAGCCCTTT-3′), *RXFP3* (KH.C6.184; 5′-CCGTGATTGTAATACGTAT GAC-3′ and 5′-GCACAGTATTGTGTTATATACC-3′), and *CNGA* (KH.C2.249; 5′-GTGGAGATGCCGATTTCAACC-3′ and 5′-CACTTTGTTTGGCATTAT TCC-3′). The *Dmrt* and *ZicL* enhancers have been described previously[16]. A similar cloning method was used to create misexpression vectors using NotI/ EcoRI sites for the *BMP2/4* (KH.C4.125; 5′-ATGGTGGCGCTTACGGATTGG AC-3′ and 5′-CTATCTACACCCACAAGCTTGC-3′) and *BMP5/7* (KH.C2.336; 5′-ATGACTTGTCATCGCAATAAAG-3′ and 5′-TCAGTTGCACCCGCATG ACAC-3′) plasmids. The *TGF-βRCA* construct was made by site-directed mutagenesis to introduce a Q154D missense mutation in the regulatory glycine- and serine-rich domain of the wild type receptor. A similar strategy was used to create the previously described *BMPR1CA* (ref. 31). The coding sequence of *SOG/ Chemokine-like* was amplified from cDNA (5′-ATGTTGCTCGGAATCAT GAAATC-3′ and 5′-GTATAAATATTCATACTTGTTTCTG-3′) and cloned into a pCESA vector using NotI/EcoRI restriction sites.

For the calcium imaging experiments, cDNA fragments containing the full-length coding regions of *C. intestinalis CNGA*, *CNGB* and *CNGC* (Ghost Gene IDs: KH.C2.249, KH.L42.6, and KH.C7.605, respectively) were amplified from a cDNA pool of mid tailbud embryos (*CNGA*) or larvae (*CNGB* and *CNGC*) by PCR using a thermostable DNA polymerase exhibiting proofreading activity (Takara LATaq; Takara Bio) with gene-specific primers (5′-AGCTAACACAGATTTTAGTGTA ATG-3′ and 5′-GTTGCGGGTAAATTACATGTC-3′ for *CNGA*; 5′-ATGGCAT TGCACATAAATGCAA-3′ and 5′-AGCAAAGACTTTGGTAAACATCAG-3′ for *CNGB*; 5′-TAAGGTTGATACAGGTTTCATTGG-3′ and 5′-GACGTAATC ATGACGAAACTGTG-3′ for *CNGC*). The PCR products were subcloned into pcDNA6 plasmid vectors and sequenced on both strands using the cycle

sequencing method with an Applied Biosystems 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

In situ hybridization. The fluorescent *in situ* hybridizations were performed as described previously[16]. The colorimetric *in situ* hybridizations on hatched larvae were performed as previously detailed[32]. The *in situ* hybridizations for Extended Data Fig. 3 were performed as specified[33]. mRNA probes were synthesized using linearized cDNA clones for *Six1/2* (ciad46p13), *Eya* (cilv29d14), *Chordin* (ciclo16e09), *BMP2/4* (cicl060n01), *GnRH2* (cilv25p08) and *RXFP3* (cinc033n18) from Nori Satoh's cDNA gene collection (OIST, Okinawa, Japan). Probe templates for *CNGA* (5′-ATGGCATTGCACATAAATGCAA-3′ and 5′-AGCAAAGAC TTTGGTAAACATCAG-3′) and *SOG/Chemokine-like* (5′-ATGTTGCTCGGA ATCATGAAATC-3′ and 5′-GTATAAATATTCATACTTGTTTCTG-3′) were amplified from a cDNA pool of mid tailbud embryos and subcloned into a pBSKM vector.

Phylogenetic analysis. For the initial survey tree, BLASTP[34] was used to assign the two *Ciona* GPCR sequences of interest to the rhodopsin class. We used the MAFFT E-INS-i algorithm[35] to align their seven-transmembrane domains with those of the proteins constituting the Pfam seed alignment for rhodopsin GPCRs[36], as well as most *Ciona* rhodopsin-receptor-like proteins identified previously[37]. After removing ambiguously aligned amino acids with Gblocks[38], the 118-sequence data set (including five outgroup sequences for *Ciona* Glutamate, Smoothened, Frizzled and cAMP receptor-like proteins) contained 373 positions (including gaps). For optimal tree searches and measurement of nodal support, minimum evolution (ME), maximum likelihood (ML) and Bayesian inference (BI) criteria were used[39].

To classify the two *Ciona* GPCRs of interest with greater accuracy and support, a second set of analyses was performed using a more focused sequence sampling. Based on the initial survey tree and the identifications made by Kamesh *et al.*[37], we selected presumed members of the specific clade containing the two *Ciona* GPCRs of interest, its large 'Chemokine Cluster' sister clade, and its two closest rootward sister clades (that is, members of the 'SOG' rhodopsin subclass). For these selections, GPCRDB Tools[40] was used to build an alignment of human seven-transmembrane domain sequences from the Swiss-Prot database. Using Clustal Omega[41], we integrated seven relevant *Ciona* sequences described by Kamesh *et al.*[37] into the GPCRDB Tools alignment. After hand-editing in Jalview[42], this 75-sequence data set contained 208 positions. As stated earlier, ME, ML and BI were used to analyse the final alignment.

Calcium imaging. HEK 293 cells, a gift from H. Yagisawa (Graduate School of Life Science, University of Hyogo), were grown at 5% CO_2 and 100% relative humidity in DMEM medium supplemented with 10% (v/v) fetal bovine serum and 0.1 mM non-essential amino acids. HEK 293 cells do not express endogenous CNG channels[43]. Each *Ciona* CNG construct was transfected into HEK 293 cells using Targefect F-1 reagent (Targeting Systems, Santee, CA, USA) according to the manufacturer's instructions, and the cells were incubated for 48 h.

To quantitatively visualize cyclic nucleotide-induced intracellular calcium ions, transfected HEK 293 cells were incubated for 30 min in a solution containing 140 mM NaCl, 5.4 mM KCl, 5 mM $CaCl_2$, 5.6 mM glucose, 10 mM HEPES (pH 7.4) and 2 μM Fura-2 AM (Dojindo Laboratories, Kumamoto, Japan). After washing with PBS, cells were placed in a solution containing 140 mM NaCl, 5.4 mM KCl, 5 mM $CaCl_2$, 5.6 mM glucose, and 10 mM HEPES (pH 7.4). 8-Br-cAMP (Sigma B7880) or 8-Br-cGMP (Sigma B1381) at various concentrations was administered to the cells in a perfusion system. Fluorescence signals were monitored using an Argus-50/CA system (Hamamatsu photonics, Shizuoka, Japan). Each experiment was performed in at least two biological replicates. Authentication and contamination tests were not performed; however, this should not affect the reliability of the results as the untransfected negative control did not show any response to cyclic nucleotides.

31. Christiaen, L., Stolfi, A. & Levine, M. BMP signaling coordinates gene expression and cell migration during precardiac mesoderm development. *Dev. Biol.* **340,** 179–187 (2010).
32. Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A. & Patel, N. H. Fixation and dissection of *Parhyale hawaiensis* embryos. *Cold Spring Harb Protoc.* **2009,** pdb.prot5127 (2009).
33. Hudson, C. & Yasuo, H. Patterning across the ascidian neural plate by lateral Nodal signalling sources. *Development* **132,** 1199–1210 (2005).
34. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
35. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
36. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–D230 (2014).
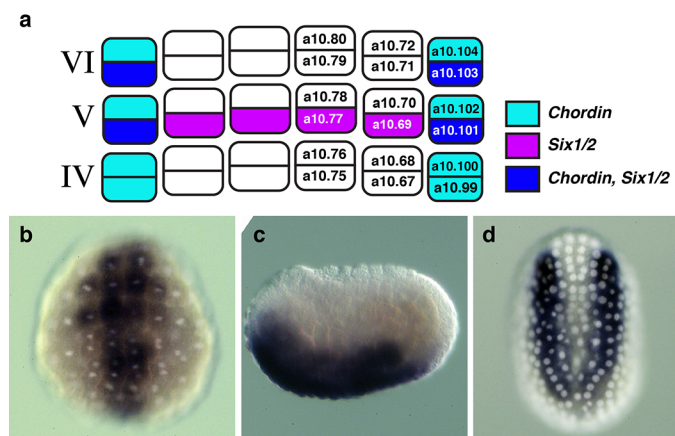
37. Kamesh, N., Aradhyam, G. K. & Manoj, N. The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. *BMC Evol. Biol.* **8,** 129 (2008).
38. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56,** 564–577 (2007).
39. Winchell, C. J. & Jacobs, D. K. Expression of the Lhx genes *apterous* and *lim1* in an errant polychaete: implications for bilaterian appendage evolution, neural development, and muscle diversification. *EvoDevo* **4,** 4 (2013).
40. Isberg, V. *et al.* GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **42,** D422–D425 (2014).
41. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
42. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
43. Dhallan, R. S., Yau, K. W., Schrader, K. A. & Reed, R. R. Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. *Nature* **347,** 184–187 (1990).
44. Hotta, K. *et al.* A web-based interactive developmental table for the ascidian *Ciona intestinalis*, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev. Dyn.* **236,** 1790–1805 (2007).
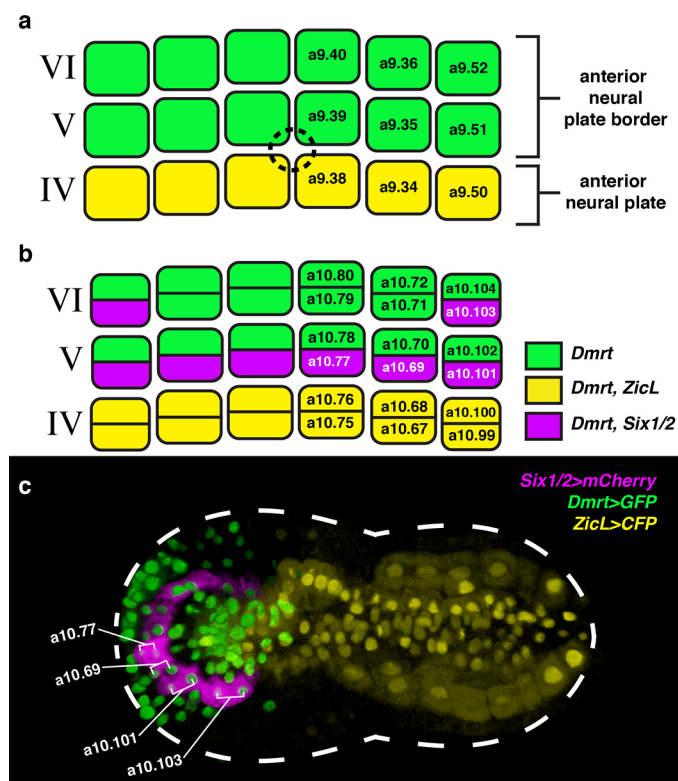
Extended Data Figure 1 | Lineage information for *Six1/2* expression in *Ciona intestinalis* from the gastrula to initial tailbud stage. **a**, Schematic of the anterior neural plate border at the mid-gastrula stage, including cell lineage nomenclature. *Dmrt* is initially activated in six blastomeres of 64-cell embryos (a7.9, a7.10 and a7.13). This lineage produces only the anterior neural plate (and the adjacent anterior neural plate border), which forms the PPE. The anterior-most ZicL[+] cells of row IV (yellow) mark the boundary of the neural plate, which gives rise to the anterior sensory vesicle in tailbud embryos. The dotted line indicates the oral opening. **b**, Schematic of the anterior neural plate border during the gastrula–neurula transition, indicating the lineage-specific expression of *Six1/2* (magenta). *Six1/2* is initially expressed in eight cells comprising the posterior cells (row V) and the posterior lateral cells (row VI). **c**, Dorsal view of an initial tailbud embryo co-electroporated with *Six1/2>mCherry*, *Dmrt>GFP* and *ZicL>CFP*. At this stage, the cells initially expressing *Six1/2* have divided once, giving rise to 16 cells in total. Brackets indicate the derivatives of the annotated lineages shown in **b**. Anterior is to the left.
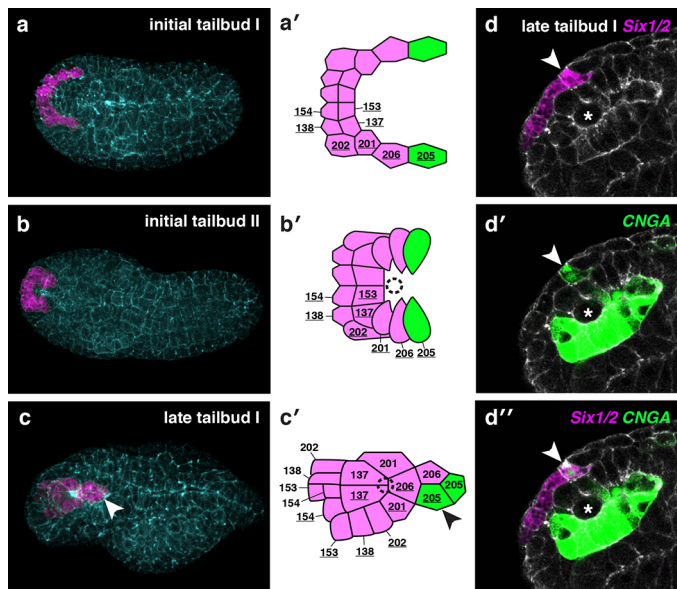
Extended Data Figure 2 | Six1/2⁺ cell morphogenesis in *Ciona intestinalis* from the initial tailbud stage to late tailbud I. a–c, Dorsal views of tailbud embryos electroporated with *Six1/2>mCherry* (magenta) and counterstained with phalloidin (blue). **a,** At the initial tailbud stage I (according to Hotta *et al.*[44]), the Six1/2⁺ cells are arranged in a U shape, anterior to the neural plate. **a′–c′,** Underlined cell lineages are derived from the left side of the embryo. **a′,** Schematic indicates that the U shape is composed of 16 cells in total at the 11th generation (that is, a11.154, a11.138, etc.). There are no further divisions of these cells until after the late tailbud I stage. The green cells indicate a11.205, which are fated to become PPE-derived GnRH neurons. **b,** At the initial tailbud stage II, the lateral edges of the Six1/2⁺ cells begin to intercalate towards the midline. **b′,** The schematic shows a dotted circle where the future opening of the oral siphon forms. **c,** At the late tailbud I stage, the Six1/2⁺ cells have completed intercalation. The bright phalloidin signal in the centre of the pattern marks the apically localized actin of cells constricted towards the oral opening. The arrowheads indicate a PPE-derived cell fated to become a GnRH neuron. At this stage, the Six1/2⁺ cells are positioned on top of the anterior sensory vesicle over the ocellus. **c′,** The schematic shows a dotted circle where the oral opening forms. More anterior cells have undergone local cellular rearrangements. **d–d″,** Anterior lateral view of late tailbud I stage embryo electroporated with *Six1/2>GFP* and *CNGA>mCherry*. Asterisks indicate the ocellus. The arrowheads indicate a PPE-derived cell fated to become a GnRH neuron. **d,** Shows the *Six1/2>GFP* channel. **d′,** Shows the *CNGA>mCherry* channel (see Fig. 3c for larval expression). **d″,** Shows the merged image of *Six1/2>GFP* and *CNGA>mCherry*.

a

| | | | | a10.80 / a10.79 | a10.72 / a10.71 | a10.104 / a10.103 |
| VI | | | | | | |

Chordin
Six1/2
Chordin, Six1/2

Extended Data Figure 3 | Dorsal–ventral BMP patterning during PPE specification in *Ciona intestinalis*. a, Schematic of the anterior neural plate border during the gastrula–neurula transition. Cell lineage nomenclature is used. *Chordin* and *Six1/2* are co-expressed in the lateral posterior derivatives of rows V and VI (dark blue; also see Fig. 1b). **b,** Ventral view of mid-gastrula stage embryo hybridized with a *BMP2/4* mRNA probe and merged with the nuclear counterstain 4′,6-diamidino-2-phenylindole (DAPI). **c,** Lateral view of an embryo during the gastrula–neurula transition hybridized with a *BMP2/4* mRNA probe. Anterior is to the left. **d,** Dorsal view of an embryo during the gastrula–neurula transition hybridized with a *Chordin* mRNA probe and merged with a DAPI nuclear counterstain.

Extended Data Figure 4 | Endogenous expression of newly described reporter genes in hatched *Ciona* larvae. a–d, Bright field anterior lateral views. **a**, Animal hybridized with a *GnRH2* mRNA probe. White arrowhead indicates the position of the oral opening. **b**, Animal hybridized with an *RXFP3* mRNA probe. **c**, Animal hybridized with a *SOG/Chemokine-like* mRNA probe. The heavily stained tunic was manually removed to reveal the expression signal. **d**, Animal hybridized with a *CNGA* mRNA probe. Red arrows mark areas of comparable expression throughout the panels in presumed PPE-derived neurons. An adjacent signal in the ocellus-associated photoreceptors makes it difficult to discriminate expression in PPE-derived neurons in panel **d**.
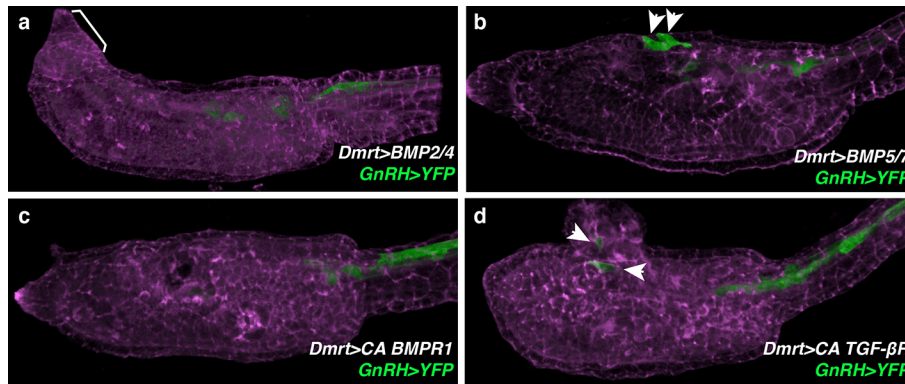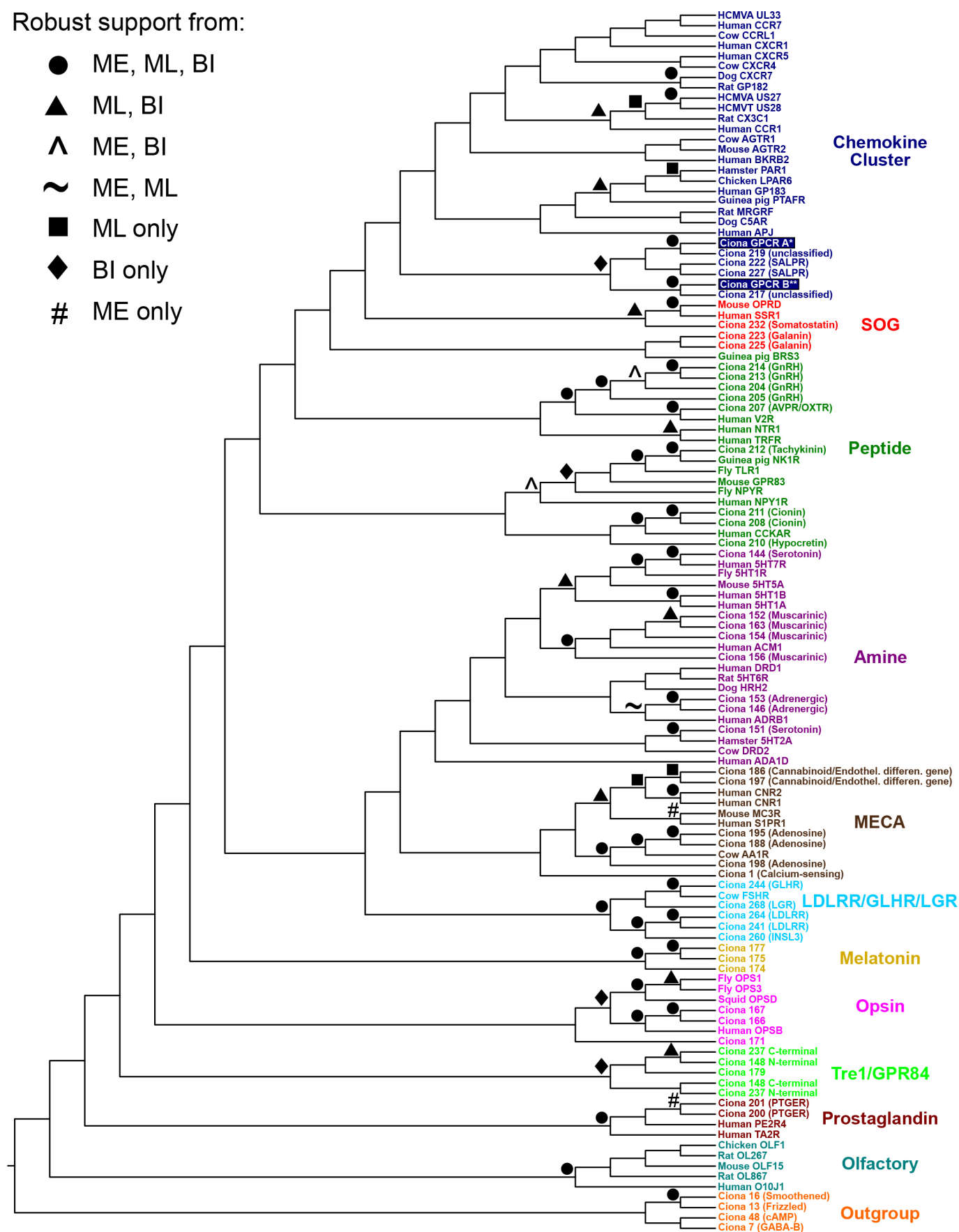
Extended Data Figure 5 | *GnRH* expression requires BMP attenuation.
a–d, Lateral view of a larva electroporated with *GnRH>YFP* and counter-stained with phalloidin (violet). **a**, Larva co-electroporated with *Dmrt>BMP2/4*. Of the 200 larvae, 196 had no *GnRH>YFP* expression in aATENs and displayed mild to severe morphogenetic defects. Bracket shows mild anterior morphological defect. **b**, Larva co-electroporated with *Dmrt>BMP5/7*. Of the 200 larvae, 124 had *GnRH>YFP* expression in aATENs. **c**, Larva co-electroporated with *Dmrt>BMPR1^{CA}*. Of the 200 larvae, 197 larvae had no *GnRH>YFP* expression in aATENs. **d**, Larva co-electroporated with *Dmrt>TGF-βR^{CA}*. Of the 200 larvae, 130 had *GnRH>YFP* expression in aATENs and most displayed severe anterior neural tube defects. Arrowheads in **b** and **d** indicate the position of *GnRH* expression in aATENs. Anterior is to the left in all images.

Robust support from:

- ● ME, ML, BI
- ▲ ML, BI
- ∧ ME, BI
- ~ ME, ML
- ■ ML only
- ◆ BI only
- # ME only

Chemokine Cluster

HCMVA UL33
Human CCR7
Cow CCRL1
Human CXCR1
Human CXCR5
Cow CXCR4
Dog CXCR7
Rat GP182
HCMVA US27
HCMVT US28
Rat CX3C1
Human CCR1
Cow AGTR1
Mouse AGTR2
Human BKRB2
Hamster PAR1
Chicken LPAR6
Human GP183
Guinea pig PTAFR
Rat MRGRF
Dog C5AR
Human APJ
Ciona GPCR A*
Ciona 219 (unclassified)
Ciona 222 (SALPR)
Ciona 227 (SALPR)
Ciona GPCR B**
Ciona 217 (unclassified)

SOG

Mouse OPRD
Human SSR1
Ciona 232 (Somatostatin)
Ciona 223 (Galanin)
Ciona 225 (Galanin)

Peptide

Guinea pig BRS3
Ciona 214 (GnRH)
Ciona 213 (GnRH)
Ciona 204 (GnRH)
Ciona 205 (GnRH)
Ciona 207 (AVPR/OXTR)
Human V2R
Human NTR1
Human TRFR
Ciona 212 (Tachykinin)
Guinea pig NK1R
Fly TLR1
Mouse GPR83
Fly NPYR
Human NPY1R
Ciona 211 (Cionin)
Ciona 208 (Cionin)
Human CCKAR
Ciona 210 (Hypocretin)

Amine

Ciona 144 (Serotonin)
Human 5HT7R
Fly 5HT1R
Mouse 5HT5A
Human 5HT1B
Human 5HT1A
Ciona 152 (Muscarinic)
Ciona 163 (Muscarinic)
Ciona 154 (Muscarinic)
Human ACM1
Ciona 156 (Muscarinic)
Human DRD1
Rat 5HT6R
Dog HRH2
Ciona 153 (Adrenergic)
Ciona 146 (Adrenergic)
Human ADRB1
Ciona 151 (Serotonin)
Hamster 5HT2A
Cow DRD2
Human ADA1D

MECA

Ciona 186 (Cannabinoid/Endothel. differen. gene)
Ciona 197 (Cannabinoid/Endothel. differen. gene)
Human CNR2
Human CNR1
Mouse MC3R
Human S1PR1
Ciona 195 (Adenosine)
Ciona 188 (Adenosine)
Cow AA1R
Ciona 198 (Adenosine)
Ciona 1 (Calcium-sensing)

LDLRR/GLHR/LGR

Ciona 244 (GLHR)
Cow FSHR
Ciona 268 (LGR)
Ciona 264 (LDLRR)
Ciona 241 (LDLRR)
Ciona 260 (INSL3)

Melatonin

Ciona 177
Ciona 175
Ciona 174

Opsin

Fly OPS1
Fly OPS3
Squid OPSD
Ciona 167
Ciona 166
Human OPSB
Ciona 171

Tre1/GPR84

Ciona 237 C-terminal
Ciona 148 N-terminal
Ciona 179
Ciona 148 C-terminal
Ciona 237 N-terminal

Prostaglandin

Ciona 201 (PTGER)
Ciona 200 (PTGER)
Human PE2R4
Human TA2R

Olfactory

Chicken OLF1
Rat OL267
Mouse OLF15
Rat OL867
Human O10J1

Outgroup

Ciona 16 (Smoothened)
Ciona 13 (Frizzled)
Ciona 48 (cAMP)
Ciona 7 (GABA-B)

Extended Data Figure 6 | Initial phylogenetic analysis of two GPCRs expressed in the PPE-derived neurons of *Ciona intestinalis*. This broad survey tree, constructed according to maximum likelihood, shows the approximate placement of the two GPCRs of interest (boxed in blue) within the rhodopsin-class G-protein-coupled receptors. The numbered *Ciona* sequences and their receptor-type identifications in parentheses are from Kamesh *et al.*[37]. All non-*Ciona* sequences were downloaded in bulk from http://pfam.xfam.org; they comprise the Pfam 'seed' alignment of seven-transmembrane receptors for the GPCRs. The rhodopsin subclass names are given to the right of each coloured group. The asterisk and double asterisk indicate the genes labelled *RXFP3* and *SOG/Chemokine-like*, respectively, in Fig. 2e. We judge robust nodal support as bootstrap percentages >70 for minimum evolution (ME) and maximum likelihood (ML) and posterior probability percentages >95 for Bayesian inference (BI).
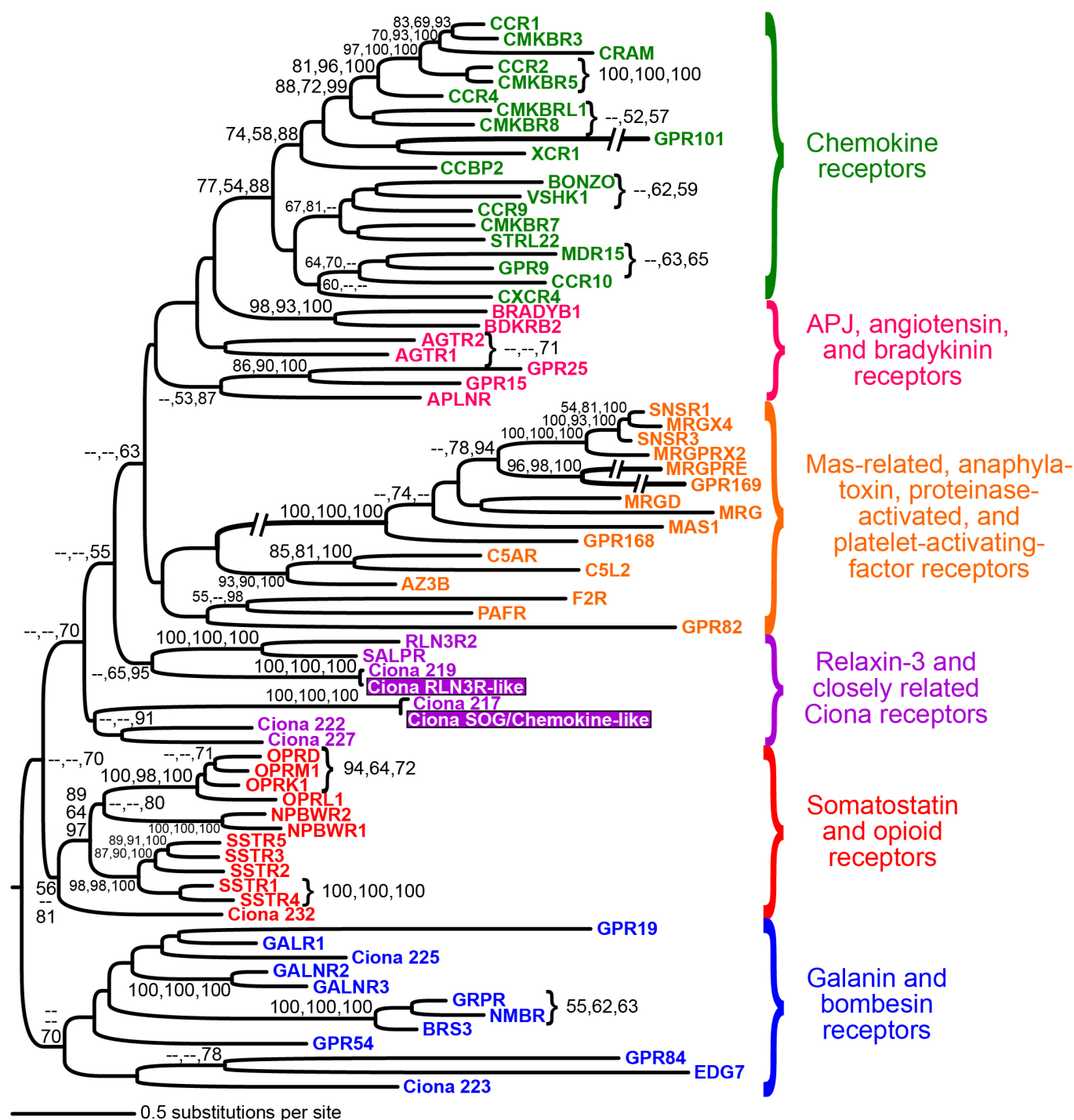
Extended Data Figure 7 | Refined maximum likelihood phylogeny of two GPCRs expressed in the PPE-derived neurons of *Ciona intestinalis.* Relative to the initial survey tree (see Extended Data Fig. 6), analysis of this focused set of 'chemokine cluster' sequences further clarifies the phylogenetic affinities of two *Ciona* GPCRs expressed in the aATEN–GnRH neurons. The latter sequences are highlighted with violet boxes; all other *Ciona* sequences are from Kamesh *et al.*[37]. The three nodal support values are (in order): ME bootstrap percentage, ML bootstrap percentage and BI posterior probability (only values >50 are shown). Branch lengths are proportional to molecular change (amino acid substitutions per site) between nodes; see scale bar for measurement.
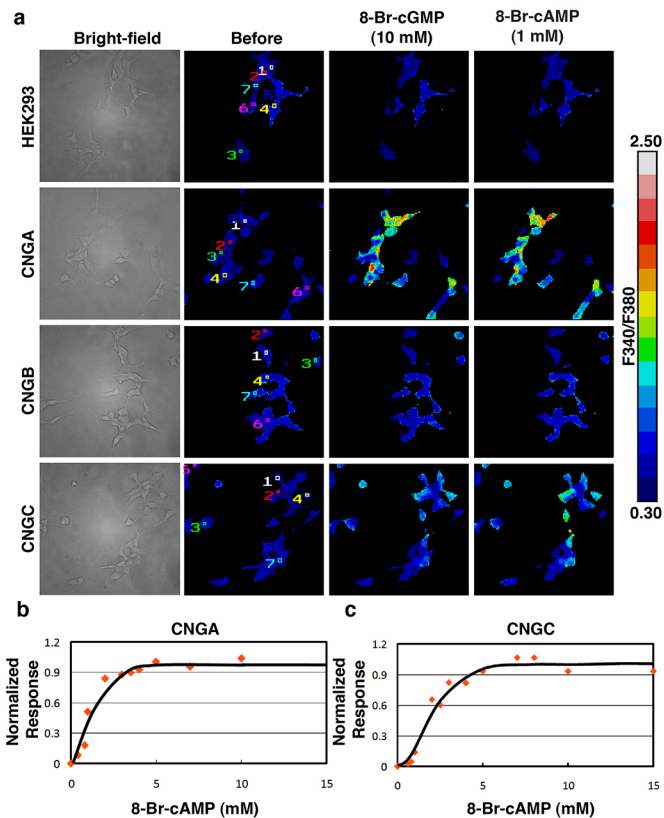
Extended Data Figure 8 | Elevation of intracellular Ca^{2+} concentration in cells expressing *Ciona* CNG channels in response to cyclic nucleotides.
a, Intracellular Ca^{2+} was visualized by Fura-2 ratiometric calcium imaging. Ca^{2+} influx into HEK 293 cells transfected with *CNGA* and *CNGC* was induced by 10 mM 8-Br-cGMP and 1 mM 8-Br-cAMP. In contrast, no Ca^{2+} influx was observed when HEK 293 cells transfected with *CNGB* were treated with either 10 mM 8-Br-cGMP or 1 mM 8-Br-cAMP. Coloured numbers in the 'Before' panels indicate cells that were subjected to a quantitative measurement of fluorescence. **b**, Dose-dependent response of HEK 293 cells transfected with *CNGA* to 8-Br-cAMP. Data were obtained from at least seven different cells in each of 16 different transfections. **c**, Dose-dependent response of HEK 293 cells transfected with *CNGC* to 8-Br-cAMP. Data were obtained from at least seven different cells in each of seven different transfections.

LETTER

Sidekick 2 directs formation of a retinal circuit that detects differential motion

Arjun Krishnaswamy[1]*, Masahito Yamagata[1]*, Xin Duan[1], Y. Kate Hong[1]† & Joshua R. Sanes[1]

In the mammalian retina, processes of approximately 70 types of interneurons form specific synapses on roughly 30 types of retinal ganglion cells (RGCs) in a neuropil called the inner plexiform layer. Each RGC type extracts salient features from visual input, which are sent deeper into the brain for further processing[1–4]. The specificity and stereotypy of synapses formed in the inner plexiform layer account for the feature-detecting ability of RGCs. Here we analyse the development and function of synapses on one mouse RGC type, called the W3B-RGC[5,6]. These cells have the remarkable property of responding when the timing of the movement of a small object differs from that of the background, but not when they coincide[6]. Such cells, known as local edge detectors or object motion sensors, can distinguish moving objects from a visual scene that is also moving[6–12]. We show that W3B-RGCs receive strong and selective input from an unusual excitatory amacrine cell type known as VG3-AC (vesicular glutamate transporter 3). Both W3B-RGCs and VG3-ACs express the immunoglobulin superfamily recognition molecule sidekick 2 (Sdk2)[13,14], and both loss- and gain-of-function studies indicate that Sdk2-dependent homophilic interactions are necessary for the selectivity of the connection. The Sdk2-specified synapse is essential for visual responses of W3B-RGCs: whereas bipolar cells relay visual input directly to most RGCs, the W3B-RGCs receive much of their input indirectly, via the VG3-ACs. This non-canonical circuit introduces a delay into the pathway from photoreceptors in the centre of the receptive field to W3B-RGCs, which could improve their ability to judge the synchrony of local and global motion.

In situ hybridization revealed that both *Sdk1* and *Sdk2* were expressed in subsets of mouse retinal neurons (Fig. 1a–c). Sdk1- and Sdk2-positive cells were largely non-overlapping, as shown previously
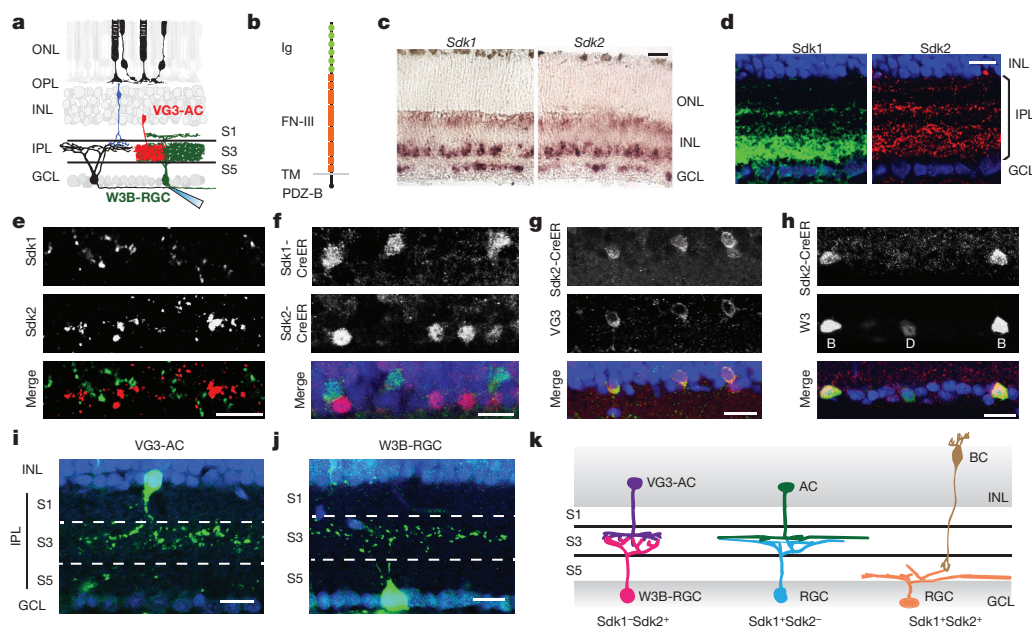

Figure 1 | Sidekick proteins are expressed by subsets of retinal neurons.
a, Photoreceptors in the outer nuclear layer (ONL) synapse on bipolar cells. Bipolar (blue), amacrine (red) and RGCs (black) synapse in the inner plexiform layer (IPL), which is divided into five sublaminae (S1–S5). W3B-RGCs are targeted for whole-cell recording and interneurons stimulated optogenetically. **b**, Structure of Sdk protein. FN-III, fibronectin type III domains; Ig, immunoglobulin domains; PDZ-B, PDZ-binding, synaptic-localizing sequence; TM, transmembrane domain. **c**, Expression of *Sdk1* and *Sdk2* in postnatal day (P) 8 retina assessed by *in situ* hybridization in retinal cross-sections. **d**, Immunohistochemical detection of Sdk1 and Sdk2 at P30. **e**, Double-label immunohistochemistry shows non-overlapping Sdk1- and Sdk2-rich puncta in S3. **f**, Double-staining for epitope-tagged CreER in *Sdk1^{ce/+} Sdk2^{ce/+}* mouse shows largely non-overlapping expression at P30. **g**, VG3-ACs are Sdk2-positive, as shown by immunostaining for VG3 and CreER in *Sdk2^{ce/+}* mouse. **h**, W3B-RGCs (B) are Sdk2-positive (CreER and YFP double staining in *Sdk2^{ce/+};TYW3* mouse). Dimmer W3D-RGCs (D) are Sdk2-negative. **i, j**, A VG3-AC (**i**) and a W3B-RGC (**j**) imaged in *Sdk2^{ce/+}* mated to a reporter line. **k**, Arborization pattern of principal retinal cell types that express *Sdk1* and *Sdk2*. The amacrine and bipolar cell types (AC and BC) that express Sdk1 remain to be determined. GCL, ganglion cell layer; INL, inner nuclear layer; ONL, outer nuclear layer. Scale bars, 10 μm.

in chicks[13,14]; however, a double-positive population was also present in mouse (Extended Data Fig. 1a). Expression was evident by embryonic day 17 and persisted into adulthood, spanning the periods of lamina formation and synaptogenesis (Extended Data Fig. 1b). Immunostaining showed that Sdk proteins were concentrated in the synapse-rich inner plexiform layer (IPL) (Fig. 1d), presumably owing to their carboxy-terminal synaptic localizing motif[15]. Sdk proteins were concentrated in two of five strata within the IPL, S3 and S5. Sdk1- and Sdk2-positive puncta in S3 were non-overlapping, consistent with the complementary expression pattern of the genes (Fig. 1e).

We generated mice in which a ligand-activated Cre recombinase (CreER) fused to distinct epitope tags was targeted to the first coding exon (Sdk1[ce] and Sdk2[ce]; Extended Data Fig. 2a–f). Interneurons and RGCs that expressed Sdk1 or Sdk2, identified using reporter lines or immunostaining, arborized in S3; neurons that expressed both Sdk1 and Sdk2 arborized in S5. Sdk1[−]/Sdk2[+] RGCs were W3B-RGCs, labelled by yellow fluorescent protein (YFP) in the TYW3 cell line, which we generated and characterized previously[5,6]. Another set of morphologically similar RGCs, called W3D, which are dimly labelled in the TYW3 line, expressed neither Sdk1 nor Sdk2. Most Sdk1[−]/Sdk2[+] interneurons expressed the vesicular glutamate transporter 3 (VGlut3), encoded by Slc17a8 (refs 16–18); we refer to these cells as VG3-ACs. Both W3B-RGCs and VG3-ACs extend dendrites that arborized in S3 (Fig. 1f–k and Extended Data Fig. 2g–i).

To determine whether VG3-ACs synapse on W3B-RGCs, we implemented an optogenetic strategy (Extended Data Fig. 3). We generated mice[5,17] in which VG3-ACs expressed channelrhodopsin 2 (ChR2) fused to a red fluorescent protein, and W3B-RGCs were labelled with YFP. We targeted YFP-positive W3B-RGCs in explanted retinas with patch electrodes, and activated ChR2 in VG3-ACs using two-photon stimulation. Optogenetic stimulation of individual VG3-ACs evoked reliable postsynaptic currents in W3B-RGCs (Fig. 2a, top trace). Displacement of the laser (~10 μm) so that it was within the receptive field of the RGC but no longer illuminated a ChR2-expressing cell evoked no stimulus-locked current (Fig. 2a, bottom trace). Thus, responses were due to excitation of ChR2 rather than photoreceptors. Additional physiological and pharmacological studies demonstrated that VG3-ACs formed excitatory, glutamatergic connections on RGCs (Extended Data Fig. 4a–c), consistent with recent studies of VGlut3-containing neurons in retina[19,20] and other brain areas[21].

We devised a test to determine whether the VG3-AC–W3B-RGC connection was monosynaptic. Although the light-activated ion channel ChR2 is highly calcium permeable, we found that it is insensitive to CdCl₂, a blocker of endogenous voltage-activated calcium channels in nerve terminals (Extended Data Fig. 4d–h). Thus, in the presence of CdCl₂, neurotransmitter can be released only from ChR2-positive terminals. Synaptic currents elicited by activating terminal arbors of ChR2-expressing VG3-ACs persisted in the presence of CdCl₂ (Extended Data Fig. 4i–k). Gap junction blockers had no effect on these currents, ruling out the possibility that calcium entering through ChR2 in VG3-ACs permeated gap junctions to electrically coupled glutamatergic bipolar cells, which then synapsed on W3B-RGCs (Extended Data Fig. 4b). Together, these results demonstrate that VG3-ACs form synapses directly on W3B-RGCs.

To assess convergence of VG3-ACs onto W3B-RGCs, we recorded from W3B-RGCs while stimulating 60–200 VG3-ACs within 200 μm of their somata. All VG3-ACs and W3B-RGCs separated by ≤100 μm were connected, with the strength of the connection inversely proportional to the distance between them (Fig. 2b). Because the radii of VG3-ACs and W3B-RGCs dendritic arbors are ~25 μm and ~60 μm, respectively[5,17,19,20], we conclude that VG3-ACs are functionally connected to W3B-RGCs whenever their dendrites overlap (Fig. 2c).

The strong connectivity of VG3-ACs to W3B-RGCs could be a simple consequence of the overlap of their arbors[22–24], as predicted by Peters' rule, which posits that connectivity is proportional to the proximity of pre- and poststynaptic arbors[22–25]. To test this idea, we
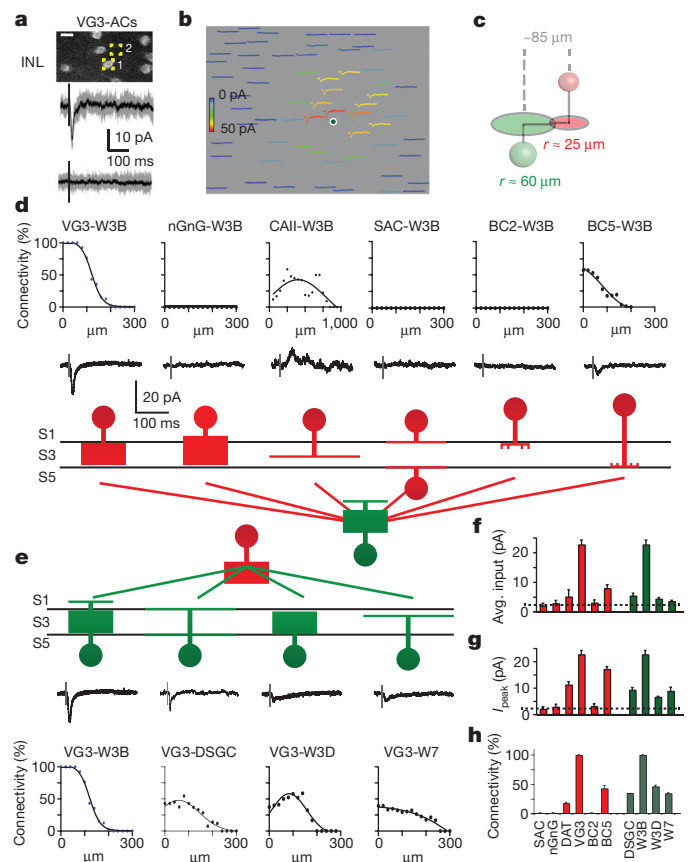

Figure 2 | Selective connectivity of VG3-ACs and W3B-RGCs. a, Current recorded from a W3B-RGC after optogenetic stimulation of a VG3-AC (top trace); no current is recorded when the stimulus is displaced by ~10 μm (bottom trace). Grey traces denote 10 trials, black traces show the average. Scale bar, 10 μm. **b,** Currents recorded in W3B-RGCs in response to stimulating 64 VG3-ACs in a 250 × 200 μm grid enclosing the RGC. Heat map indicates peak strength of connections; position of trace corresponds to position of stimulated VG3-AC soma. **c,** Cartoon showing the radii of VG3-AC and W3B-RGC arbors, indicating distance over which overlap occurs. **d, e,** Strength of connections as a function of distance from six interneuron types to W3B-RGCs (**d**) and from VG3-ACs to four RGC types (**e**). Sample currents shown below each graph. The W7 population contained six nearly disconnected and ten connected pairs, presumably corresponding to the S1-laminating and S3-laminating W7 subsets. Full data are shown in Extended Data Fig. 5. Upward currents are inhibitory (holding potential (V_h) ≈ 0 mV); downward currents are excitatory (V_h ≈ −60 mV). BC2, type 2 bipolar cells; BC5, type 5 bipolar cells; CAII, type II catecholaminergic cells. **f,** Average strength of connectivity for pairs shown in **d**. **g,** Strength of connections (average peak current, I_{peak}) for pairs in which responses exceeded baseline (dotted line, average noise floor, 2–3 pA). V_h = −70 mV for inward currents, and −10 mV for outward currents. **h,** Percentage of connected pairs for VG3-AC to RGC recordings and interneuron to W3B-RGC recordings averaged over 100 μm radius from the postsynaptic RGC. Full data and n values for **f–h** are shown in Extended Data Fig. 5. Error bars in **f–h** denote mean ± s.e.m.

measured the connectivity of VG3-ACs to four Sdk2-negative RGC types for which we had marker lines: W3D-RGCs, W7-RGCs and two types of ON–OFF direction-selective RGCs (ooDSGCs); W3D-RGC and W7-RGC dendrites intermingle with those of W3B-RGCs, and ooDSGC dendrites straddle those of W3B-RGCs. Optogenetic stimulation of VG3-ACs elicited excitatory postsynaptic currents from W3D-RGCs, W7-RGCs and ooDSGCs that were qualitatively similar to, but only ~10% as strong as those in W3B-RGCs (Fig. 2e, f and Extended Data Fig. 5). The weakness resulted from a twofold decrease in the fraction of pairs that were detectably connected, and a fivefold decrease in the peak synaptic currents for connected pairs (Fig. 2g, h).

We also assayed input to W3B-RGCs from six other types of inter-neurons that arborize in or at the border of S3. In all cases, connectivity was many-fold lower than that between VG3-ACs and W3B-RGCs (Fig. 2d–h). Together, these data demonstrate that Peters' rule is insufficient to explain patterns of connectivity in the neuropil of the retina.

To address whether Sdk2 is involved in the establishment of the strong VG3-AC to W3B-RGC connection, we used the *Sdk2^ce* line, in which insertion of CreER generates a *Sdk2*-null allele (Extended Data Fig. 2). *Sdk2^{ce/ce}* mice are viable, fertile and exhibit no external deficits. We detected no alterations in retinal structure or in the numbers or positions of any cell types examined (Extended Data Fig. 6). However, physiological analysis revealed a 20-fold reduction in the strength of synaptic connections between VG3-ACs and W3B-RGCs (Fig. 3a–d). Thus, Sdk2 is required for the selective connectivity of VG3-ACs to W3B-RGCs.

We sought morphological correlates of the synaptic disruption observed in Sdk2 mutants. To this end, we imaged single VG3-ACs and W3B-RGCs labelled in a transgenic line or by dye injection. The size, shape and laminar restriction of VG3-AC and W3B-RGC arbors were generally normal in *Sdk2^{ce/ce}* mice (Fig. 3e, f). However, branch number and length were modestly reduced in mutant W3B-RGCs and modestly increased in VG3-ACs; in both cell types, some branches extended beyond the normal termination zone (Fig. 3g–l and Extended Data Figs 6m–o and 7). Defects in laminar restriction were similar to, although less striking than, those observed in chick retina after microRNA-mediated attenuation of Sdk2 (ref. 13). In chicks, *SDK2* expression was decreased in isolated cells in a wild-type background; we speculate that this led to competition between normal and mutant neurons resulting in greater defects than observed in the null mouse mutant.

The result that Sdk2-positive neurons connect more strongly to each other than to Sdk2-negative partners is consistent with the finding that Sdk proteins are homophilic adhesion molecules[12,13]. If this is true, defects should be observed only at synapses in which both partners express *Sdk2*. To test this idea, we asked whether *Sdk2* deletion affected synapses in which only one partner was Sdk2-positive. Deletion of *Sdk2* had no effect on the strength of coupling in any of four such cases: connections of *Sdk2*-negative bipolar and amacrine cells to W3B-RGCs and of VG3-ACs to *Sdk2*-negative RGCs (Fig. 3b–d and Extended Data Fig. 8a–r). These results also provide evidence that loss of Sdk2 does not affect the overall properties of VG3-ACs or W3B-RGCs. In addition, we generated mice in which *Sdk2* could be expressed in a Cre-dependent manner in any cell (Extended Data Fig. 8s). In combination with the *Sdk2^{ce/ce}* line, this allowed us to restore Sdk2 in VG3-ACs selectively. In this case, connectivity was as low as in the *Sdk2^{ce/ce}* mutants (Extended Data Fig. 8t). Together, these results support the idea that Sdk2 promotes connectivity by a homophilic mechanism.

We next asked what role the Sdk2-specified VG3-AC-to-W3B-RGC connection has in the function of W3B-RGCs. To this end, we recorded responses of W3B-RGCs to visual rather than optogenetic stimulation. As reported previously[5,6,20], a bright spot flashed over the dendritic arbor (the receptive field centre) of control W3B-RGCs elicited a burst of action potentials at both the onset and the offset of the light, with the OFF response substantially larger than the ON response. In *Sdk2^{ce/ce}* mice, the ON response persisted but the OFF response was nearly abolished (Fig. 4a and Extended Data Fig. 9a). This result was unexpected, because input from amacrine cells generally modulates RGC responses but does not generate them[3]. To address whether this phenotype resulted from decreased excitation or enhanced inhibition, we recorded synaptic currents in response to the same stimulus. Consistent with results from the voltage recording, a flashing spot elicited excitatory currents at both light onset and offset, with OFF currents larger than ON currents; in *Sdk2* mutants, the OFF current was nearly abolished while the ON current was less affected (Fig. 4c, d). By contrast, inhibitory currents evoked in W3B-RGCs by full-field
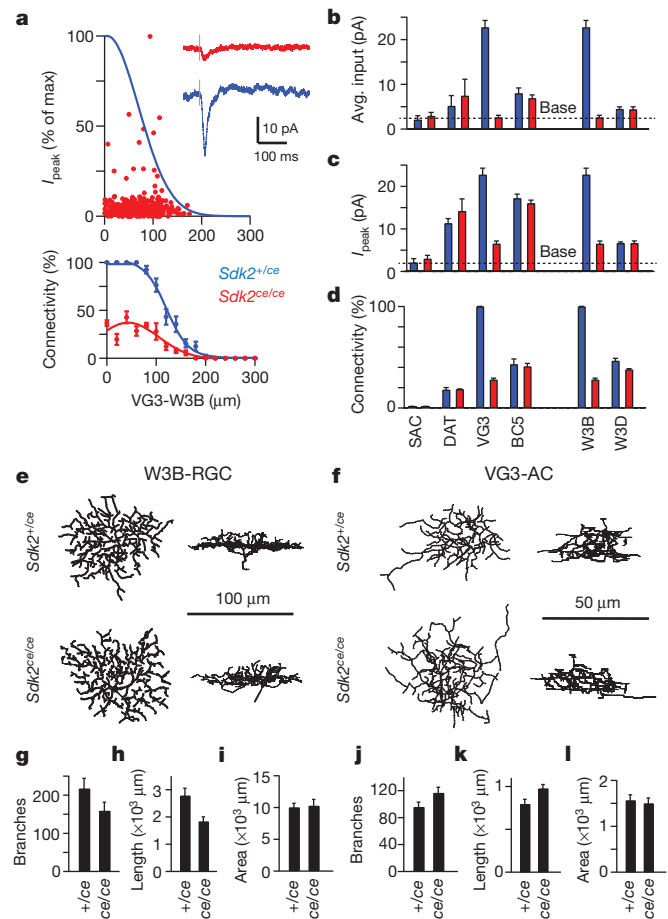

Figure 3 | Decreased synaptic connectivity in *Sdk2* mutants. a, Strength of connections from VG3-ACs to W3B-RGCs in *Sdk2^{ce/ce}* mutants, derived from 557 VG3-ACs and 12 W3B-RGCs. Fit to the control data (Extended Data Fig. 5a) re-plotted in blue. Inset, VG3-AC-evoked currents recorded from a W3B-RGC in *Sdk2^{ce/ce}* and *Sdk2^{ce/+}* retinae. **b–d**, Synaptic connectivity in *Sdk2^{ce/ce}* mutants and those in controls. Graphs show total strength of connections (**b**), strength of connections for pairs in which responses exceeded baseline (**c**, dotted line) and percentage of connected pairs (**d**). Full data are shown in Extended Data Fig. 8. **e, f**, En face and laminar projections of skeletonized dendritic arbors from dye-filled W3B-RGCs (**e**) or VG3-ACs labelled sparsely with tdTomato in *Sdk2^{+/ce}* and *Sdk2^{ce/ce}* retinae (**f**). **g–i**, Average branch number, branch length, and dendritic field area for W3B-RGCs in *Sdk2^{ce/ce}* and *Sdk2^{ce/+}* retinae. Data from 14 W3B-RGCs in 8 *Sdk2^{+/ce}* mice and 11 W3B-RGCs in 9 *Sdk2^{ce/ce}* mice. **j–l**, Similar to **g–i**, except for 9 VG3-ACs in 6 *Sdk2^{+/ce}* mice and 10 VG3-ACs in 8 *Sdk2^{ce/ce}* mice. Data in **b–d** and **g–l** indicate mean ± s.e.m.

stimulation, presumably derived from conventional amacrine cells, were unaffected in *Sdk2* mutants (Fig. 4e and Extended Data Fig. 9b). The effect was specific in that light-evoked inward currents in neighbouring RGCs, such as W3D-RGCs, persisted in *Sdk2* mutants (Extended Data Fig. 9c).

The effect we observed could have resulted from defects in other *Sdk2*-expressing retinal cells (Fig. 1k), or compensatory alterations during development. To test these possibilities, we selectively ablated mature VG3-ACs. We expressed the diphtheria toxin receptor (DTR) in VG3-ACs, injected diphtheria toxin in adult animals, recorded from W3B-RGCs ~10 days after injection, and verified loss of VG3-ACs after recording (Extended Data Fig. 9d, e). Ablation of VG3-ACs in adulthood led to an even greater loss of light-evoked excitatory OFF responses than observed after global deletion of *Sdk2*; inhibitory responses in W3B-RGCs and excitatory responses in nearby RGCs were unaffected (Fig. 4e, g and Extended Data Fig. 9f–h; see also ref. 20). Together, these results lead to the conclusion that visual input are
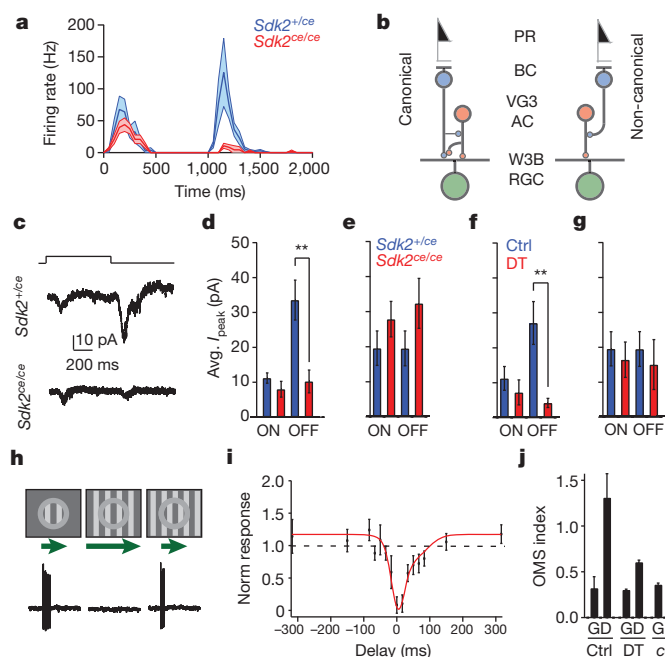
Figure 4 | Excitatory input reaches W3B-RGCs via VG3-ACs. a, Average firing rate recorded from W3B-RGCs in $Sdk2^{ce/+}$ (blue, $n = 21$) and $Sdk2^{ce/ce}$ (red, $n = 10$) retinae in response to a small spot flashed for 1 s. Dark lines, average; shadowing denotes s.e.m. Bin width, 50 ms. **b**, Canonical and non-canonical pathways for delivering OFF input to RGCs. **c**, Excitatory currents recorded from W3B-RGCs in response to a spot flash ($V_h = -65$ mV). **d**, Average peak current (I_{peak}) from experiments such as those shown in **c** ($n = 7$ W3B-RGCs in 4 $Sdk2^{ce/+}$ mice and 8 W3B-RGCs in 4 $Sdk2^{ce/ce}$ mice; **$P < 0.01$, Student's t-test). **e**, I_{peak} of inhibitory currents recorded from W3B-RGCs in response to a full field flash ($V_h = -5$ mV) ($n = 7$ W3B-RGCs in 5 controls and 5 W3B-RGCs in 4 $Sdk2^{ce/ce}$ mice). **f**, I_{peak} of excitatory currents recorded from W3B-RGCs in response to a spot flash ($n = 5$ W3B-RGCs in 3 control mice and 11 W3B-RGCs in 5 $Vglut3$-$cre;DTR$ mice; **$P < 0.01$, Student's t-test). **g**, I_{peak} of inhibitory currents recorded from W3B-RGCs in response to a full field flash ($n = 5$ W3B-RGCs in 5 mice (4 $Sdk2^{ce/+}$ mice and 1 control mouse pooled) and 5 diphtheria toxin (DT)-treated $Vglut3$-$cre;DTR/TYW3$ mice). **h**, Object motion stimulus-evoked responses in W3B-RGCs. W3B-RGCs spike vigorously to a grating passed over their receptive field centre (local). They are silenced when the moving grating extends to their surround (global) but not when the centre grating leads or lags that in the surround (differential), as quantified in **i. i**, Responses measured in W3B in response to a stimulus where the centre grating began its movement at different times relative to that of the surround. Negative delays are when the centre led the surround. Responses were normalized (Norm) to those evoked by a centre grating alone ($n = 7$ W3B-RGCs). **j**, Responses measured in W3B-RGCs to global (G) and differential (D) motion stimuli normalized to that elicited in control W3B-RGCs with local motion stimuli ($n = 10$ W3B-RGCs in 5 control mice, 6 W3B-RGCs in 4 $Sdk2^{ce/ce}$ mice and 9 W3B-RGCs in 3 diphtheria-toxin-treated $Vglut3$-$cre;DTR/TYW3$ mice). OMS, object motion sensitivity. Data in **e–g**, **i** and **j** are mean ± s.e.m.

delivered to W3-RGCs in an unusual way: while bipolar cells synapse directly on most RGCs[3], W3B-RGCs receive OFF input indirectly via VG3-ACs (Fig. 4b). The small, statistically insignificant decrease in the ON response may reflect the presence of other ON inputs that compensate for loss of that normally supplied by VG3-ACs.

Speed is of the essence for visual perception, so it seems odd to interpose an extra synapse between photoreceptors and W3B-RGCs. A clue to the reason for this comes from the fact that W3B-RGCs compare motion in the centre and surround of the receptive field, firing only when the two are asynchronous[5–7] (Fig. 4h, i). For the comparison to be temporally precise, input from the surround must arrive at the cell rapidly and/or input from the centre must be delayed. Previous studies[5,7] show that spiking amacrine cells convey rapid inhibition to object-motion-sensing RGCs including W3B-RGCs.

We find that in addition, currents elicited by stimulation of the receptive field centre are delayed in W3B-RGCs relative to other RGCs (Extended Data Fig. 10). We speculate that the interposition of VG3-ACs between bipolar cells and W3B-RGCs contributes to this delay, although it cannot explain it entirely. In addition, VG3-ACs themselves are tuned to differential motion[20], presumably accounting for much of the tuned excitatory input that W3B-RGCs receive[6]. Consistent with this interpretation, the ability of W3B-RGCs to distinguish local from global motion is markedly reduced in Sdk2 mutants or after elimination of VG3-ACs, as evidenced by a reduction in spike rate relative to control for local and differential motion (Fig. 4j and Extended Data Fig. 10i).

In summary, the recognition molecule Sdk2 is required for the selective and strong connectivity between Sdk2-positive VG3-ACs and W3B-RGCs. Connections of VG3-ACs or W3B-RGCs with proximate but Sdk2-negative partners are substantially weaker. Because Sdk2 is localized at synapses and required in both partners, we speculate that it acts homophilically to promote appropriate connections. In its absence, functional connections fail to form or are not maintained, leading to markedly decreased synaptic strength.

Taken together with previous studies, our data support a multi-step model for synaptic specificity in the IPL. First, one set of recognition molecules, including cadherins and plexins, direct arbors to appropriate sublaminae[26,27]. Within sublaminae, proximate partners connect at low levels, consistent with Peters' rule[22–25]. Finally, recognition molecules such as Sdk proteins, and perhaps other immunoglobulin superfamily adhesion molecules[28,29], act to bias connectivity in favour of specific pairings.

Our results also reveal a role of the VG3-AC–W3B-RGC synapse in visual function. In canonical retinal circuits, bipolar cells relay visual input from photoreceptors to RGCs, whereas amacrine cells, which have been presumed to be inhibitory, modulate this input[3] (Fig. 4b). By contrast, VG3-ACs provide the main excitatory drive to the W3B-RGCs. Thus, the VG3-AC–W3B-RGC synapse is a component of a non-canonical retinal circuit in which some of the visual input is relayed to W3B-RGCs through VG3-ACs rather than arriving directly through bipolar cells. This seemingly cumbersome arrangement could improve the sensitivity of W3B-RGCs to the visual features that best excite them—the motion of small objects whose movements are asynchronous with those of the background[5].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Sanes, J. R. & Zipursky, S. L. Design principles of insect and vertebrate visual systems. *Neuron* **66,** 15–36 (2010).
2. Gollisch, T. & Meister, M. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* **65,** 150–164 (2010).
3. Masland, R. H. The neuronal organization of the retina. *Neuron* **76,** 266–280 (2012).
4. Sanes, J. R. & Masland, R. H. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annu. Rev. Neurosci.* **38,** 221–246 (2015).
5. Kim, I. J., Zhang, Y., Meister, M. & Sanes, J. R. Laminar restriction of retinal ganglion cell dendrites and axons: subtype-specific developmental patterns revealed with transgenic markers. *J. Neurosci.* **30,** 1452–1462 (2010).
6. Zhang, Y., Kim, I. J., Sanes, J. R. & Meister, M. The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proc. Natl Acad. Sci. USA* **109,** E2391–E2398 (2012).
7. Baccus, S. A., Olveczky, B. P., Manu, M. & Meister, M. A retinal circuit that computes object motion. *J. Neurosci.* **28,** 6807–6817 (2008).
8. Ölveczky, B. P., Baccus, S. A. & Meister, M. Segregation of object and background motion in the retina. *Nature* **423,** 401–408 (2003).
9. Venkataramani, S. *et al.* Distinct roles for inhibition in spatial and temporal tuning of local edge detectors in the rabbit retina. *PLoS ONE* **9,** e88560 (2014).
10. Russell, T. L. & Werblin, F. S. Retinal synaptic pathways underlying the response of the rabbit local edge detector. *J. Neurophysiol.* **103,** 2757–2769 (2010).
11. van Wyk, M., Taylor, W. R. & Vaney, D. I. Local edge detectors: a substrate for fine spatial vision at low temporal frequencies in rabbit retina. *J. Neurosci.* **26,** 13250–13263 (2006).

12. Levick, W. R. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *J. Physiol. (Lond.)* **188,** 285–307 (1967).
13. Yamagata, M., Weiner, J. A. & Sanes, J. R. Sidekicks: synaptic adhesion molecules that promote lamina-specific connectivity in the retina. *Cell* **110,** 649–660 (2002).
14. Yamagata, M. & Sanes, J. R. Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. *Nature* **451,** 465–469 (2008).
15. Yamagata, M. & Sanes, J. R. Synaptic localization and function of Sidekick recognition molecules require MAGI scaffolding proteins. *J. Neurosci.* **30,** 3579–3588 (2010).
16. Haverkamp, S. & Wassle, H. Characterization of an amacrine cell type of the mammalian retina immunoreactive for vesicular glutamate transporter 3. *J. Comp. Neurol.* **468,** 251–263 (2004).
17. Grimes, W. N., Seal, R. P., Oesch, N., Edwards, R. H. & Diamond, J. S. Genetic targeting and physiological features of VGLUT3$^+$ amacrine cells. *Vis. Neurosci.* **28,** 381–392 (2011).
18. Johnson, J. *et al.* Vesicular glutamate transporter 3 expression identifies glutamatergic amacrine cells in the rodent retina. *J. Comp. Neurol.* **477,** 386–398 (2004).
19. Lee, S. *et al.* An unconventional glutamatergic circuit in the retina formed by vGluT3 amacrine cells. *Neuron* **84,** 708–715 (2014).
20. Kim, T., Soto F. & Kerschensteiner D. An excitatory amacrine cell detects object motion and provides feature-selective input to ganglion cells in the mouse retina. *eLife* http://dx.doi.org/10.7554/eLife.08025 (2015).
21. El Mestikawy, S. *et al.* From glutamate co-release to vesicular synergy: vesicular glutamate transporters. *Nature Rev. Neurosci.* **12,** 204–216 (2011).
22. Peters, A. & Feldman, M. L. The projection of the lateral geniculate nucleus to area 17 of the rat cerebral cortex. I. General description. *J. Neurocytol.* **5,** 63–84 (1976).
23. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* **28,** 387–394 (2005).
24. Shepherd, G. M., Stepanyants, A., Bureau, I., Chklovskii, D. & Svoboda, K. Geometric and functional organization of cortical circuits. *Nature Neurosci.* **8,** 782–790 (2005).
25. Binzegger, T., Douglas, R. J. & Martin, K. A. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24,** 8441–8453 (2004).
26. Duan, X., Krishnaswamy, A., De la Huerta, I., Sanes, J. R. & Type, I. I. Cadherins guide assembly of a direction-selective retinal circuit. *Cell* **158,** 793–807 (2014).
27. Matsuoka, R. L. *et al.* Transmembrane semaphorin signalling controls laminar stratification in the mammalian retina. *Nature* **470,** 259–263 (2011).
28. Yamagata, M. & Sanes, J. R. Expanding the Ig superfamily code for laminar specificity in retina: expression and role of contactins. *J. Neurosci.* **32,** 14402–14414 (2012).
29. Fuerst, P. G., Koizumi, A., Masland, R. H. & Burgess, R. W. Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature* **451,** 470–474 (2008).

METHODS

Animals. We modified a lambda-phage-mediated recombineering method[30] to generate *Sdk1* and *Sdk2* targeting vectors in which the first coding exons of the *Sdk1* and *Sdk2* genes were replaced by an epitope-tagged CreER-T2 recombinase complementary DNA. Tags were three tandem copies of the haemagglutinin (HA) tag (amino acid sequence YPYDVPDYA) for Sdk1, and six tandem copies of the Myc tag (amino acid sequence EQKLISEEDL) for Sdk2. Loci were modified by homologous combination in V6.5 embryonic stem cells, and chimaeras were produced by the Harvard University Genome Modification Facility. High percentage chimaeras transmitting the knock-in alleles were bred to animals expressing FLP recombinase from the β-actin promoter[31] to remove the SV40-NEO cassette.

Thy1-STOP-YFP, *TYW3*, *TYW7*, *TYW9*, *Kcng4-cre* and *Neto1-cre* mice were generated in our laboratory. *Thy1-STOP-YFP* expresses YFP after excision of a stop cassette by Cre recombinase[32]. *TYW3* mice express YFP in subsets of RGCs[5]. *TYW7* mice express YFP in two types of RGCs, which are present at similar densities; one has its dendrites in the distal half of S3; the other has its dendrites in S1. *TYW9* mice express YFP in ooDSGCs that prefer nasal motion[33]. *Neto1-cre* and *Kcng4-cre* mice express Cre recombinase in type 2 and type 5 bipolar cells, respectively, as well as in subsets of RGCs[26,34]. *Vglut3-cre* mice were obtained from R. Seals and R. Edwards. In these mice, Cre was inserted into exon 1 of a bacterial artificial chromosome that contained *Vglut3*, and this vector was used to generate a transgenic line[11]. *Nex-cre* mice, in which Cre is targeted to the endogenous *Neurod6* locus[35], were obtained from K. Nave via L. Reichardt; this line expresses Cre in non-GABAergic non-glycinergic (nGnG) and glycinergic SEG amacrine cells[36]. *DAT-cre* mice, in which Cre is targeted to the endogenous *DAT* (also known as *Slc6a3*) locus, were obtained from X. Zhuang[37] via V. Murthy; this line expresses Cre in type II catecholaminergic amacrine cells. *ChAT-cre* mice, in which the Cre recombinase gene was targeted to the endogenous *Chat* gene, was obtained from Jackson Laboratories[38]; this line expresses Cre in starburst amacrine cells (SACs). Hb9-GFP transgenic mice, which express green fluorescent protein (GFP) in ooDSGCs that prefer nasal motion[39], were obtained from K. Eggan. *Rosa-LOX-STOP-LOX-ChR2(H134R)-tdTomato* mice (Ai27), which express channelrhodopsin after excision of a stop cassette by Cre recombinase[40] were provided by H. Zeng. *Rosa-CAG-LOX-STOP-LOX-DTR* mice, which express DTR following excision of a stop cassette by Cre recombinase, were obtained from Jackson Laboratories[41]. The *Rosa-CAGS-LOX-CHERRY-LOX-GFP* line was from S. Dymecki[42]. *Six3-cre* mice[43] were provided by W. Klein.

To enable expression of Sdk2 under Cre-dependent control, we generated a line using a previously described strategy[44,45]. A cassette encoding Venus and *Sdk2*, separated by tripleF2A (three tandem repeats of foot-and-mouth disease 2A peptide sequence) was cloned into a *Rosa26-CAG-STOP*-targeting vector to generate *Rosa-CAG-LOX-STOP-LOX-Venus-3F2A-Sidekick2-WPRE-FRT-neo-FRT*. Homologous recombinants were selected in the V6.5 embryonic stem cell line and chimaeras were generated. Germ-line chimaeras were crossed to a Flp mouse[31] to obtain germ-line transmissions and to remove the FRT-neo-FRT sequence.

Animals were used in accordance with NIH guidelines and protocols approved by Institutional Animal Use and Care Committee at Harvard University. Mice were maintained on a C57/B6J background. Both male and female mice were used in this study. Animals were 40 to 100 days old at the time of euthanasia unless otherwise stated in the text or figure legend. Genotypes of mice were known to investigators at the time of the experiment, and there was no randomization in assignment of animals for specific experiments.

RT–PCR. RNA was prepared from the brains of wild-type or knockout mice using EZNA Total RNA kit I (OMEGA bio-tek), reverse-transcribed by SuperScript III (Life Technologies), and amplified using Taq DNA polymerase (EconoTaq PLUS, Lucigen) with these gene-specific PCR primers. *Sdk1* forward: 5′-TGAACGGTCTTCTGCAAGGCTACA-3′, *Sdk1* reverse: 5′-AAGGGTCAGCTCAGAGGCAGATTT-3′; *Sdk2* forward: 5′-TTCTGGCTGGTAGAAGGCAACTCA-3′, *Sdk2* reverse: 5′-AGGATGCCGTTGATCTTGTCCTCA-3′; *Cre* forward: 5′-GCATTACCGGTCGATGCAACGAGTGATGAG-3′, *Cre* reverse: 5′-GAGTGAACGAACCTGGTCGAAATCAGTGCG-3′; *Gapdh* forward: 5′-TGAAGGTCGGTGTGAACGGATTTGGC-3′, *Gapdh* reverse: 5′-CATGTAGGCCATGAGGTCCACCAC-3′.

Histology. Mice were euthanized by intraperitoneal injection of pentobarbital and either enucleated immediately or transcardially perfused with Ringer's solution followed by 4% (w/v) paraformaldehyde (PFA) in PBS. Eye cups were removed and fixed in 4% (w/v) PFA in PBS on ice overnight, sunk in 30% (w/v) sucrose/PBS, and mounted in the OCT compound. Immunostaining of cryosections was carried out as described previously[14,44]. For double immunostaining with two different mouse antibodies, we used the Zenon Horseradish Peroxidase Mouse IgG1 Labelling Kit (Life Technologies) to label one of them, and detected reaction product with TSA-Plus kits (Perkin-Elmer Life Sciences). For immunodetection of epitope-tagged CreER, cryosections were permeabilized in absolute methanol at

−20 °C overnight, treated with Image-iT FX signal enhancer (Life Technologies) as instructed by the manufacturer's protocol, and blocked with 5% (w/v) skim milk (BioRad) in PBS for 30 min at room temperature. The antibodies were diluted in Renoir Red diluent (BioCare Medical), incubated at 4 °C for 48 h, rinsed, and detected with secondary antibodies which had been preabsorbed with acetone powders prepared from mouse brain. In some cases, animals were injected twice with 1 mg tamoxifen (Sigma) in 0.1 ml sunflower oil at 24 and 48 h before culling, resulting in a higher concentration of CreER in the nucleus, which enhanced our ability to detect it.

To label single VG3-ACs and W3B-RGCs, retinae containing dye-filled W3B-RGCs or sparsely labelled VG3-ACs were fixed in 4% PFA at 4 °C for 1 h and then incubated in primary antibodies dissolved in blocking solution (PBS, 0.3% Triton X-100 and 3% donkey serum; Jackson Immunoresearch) for 7–10 days at 4 °C with agitation. Next, retinae were washed for 3–5 h in 3–5 changes of PBS and incubated with secondary antibodies overnight at 4 °C with agitation. After 3–5 h of washing in PBS with agitation, retinae were flattened onto nitrocellulose membranes (Millipore) and mounted on slides (Vectashield, Vector Labs).

To generate antibodies to mouse Sdk1 and Sdk2, ~100 amino-acid-long stretches of the intracellular domains were selected in a region where the two molecules maximally differ in amino acid sequence. cDNAs encoding these fragments fused to a poly-histidine tags were inserted into a pET vector (Novagen). Fusion proteins were produced in BL21 bacteria and purified using a His-column (Life Technologies). Animals were immunized and antisera produced by Covance Research. Antibodies were affinity purified using the antigen fusion proteins as bait. We also generated mouse polyclonal antibodies to mouse Sdk1 and Sdk2 by immunizing Sdk1- and Sdk2-knockout mice with L cells (ATCC) that had been transfected with full-length mouse *Sdk1* or *Sdk2* cDNA as described previously[28]. Antibodies were tested on *Sdk1*- and *Sdk2*-expressing HEK cells (HEK-293T; ATCC) and on retinal tissue from *Sdk1ce/ce* and *Sdk2ce/ce* mice to verify specificity.

Antibodies used in this study were: rabbit monoclonal antibody to oestrogen receptor alpha (Clone SP1, from Epitomics or Abcam); goat anti-Myc (NB600-335, from Novus); rat anti-HA (3F10, from Roche Diagnostics); anti-Brn3a (clone, 5A3.2), rabbit anti-synapsin I (AB1543P), mouse anti-calretinin (clone, 6B8.2), goat anti-ChAT antibodies, and sheep anti-tyrosine hydroxylase from Millipore; AP2 (clone, 3B5), SV2, anti-synaptotagmin 2 (clone, ZNP1) from Developmental Studies Hybridoma Bank; mouse anti-VGluT1 (clone, N28/9), mouse anti-pan-MAGUK (clone, N28/86), mouse anti-HCN4 (clone, N114.10), and mouse anti-Vesicular acetylcholine transporter (clone, N6/38) from NeuroMab; mouse anti-SATB2 (clone, 4B19) from Abcam; rabbit anti-fluorescein (Life Technologies); and rabbit anti-protein kinase C alpha (P4334) from Sigma. Rabbit antibody to Dab1 was a gift from B. Howell. Rabbit anti-lucifer yellow and anti-fluorescein were from Invitrogen. Chicken anti-GFP and rabbit anti-mCherry were generated as described previously[42,44]. Nuclei were labelled with NeuroTrace Nissl 435 (Life Technologies). Secondary antibodies were conjugated to DyLight 488, DyLight 594, or Alexa 647 (Jackson ImmunoResearch).

For *in situ* hybridization, riboprobes were synthesized from *Sdk1* or *Sdk2* cDNAs using digoxigenin- or fluorescein-labelled UTP and hydrolysed to around 500 base pairs as described previously[13,15]. Probes were detected using anti-digoxigenin antibodies conjugated to alkaline phosphatase, followed by reaction with BCIP (5-bromo-4-chloro-3-indolyl phosphate) and NBT (nitroblue tretrazolium) substrate for 24–36 h; or using anti-digoxigenin and anti-fluorescein antibodies conjugated to horseradish peroxidase, followed by amplification with tyramide conjugates (TSA-Plus system; Perkin-Elmer Life Sciences).

Imaging. Images of immunostained retinal wholemounts were acquired on a LSM 710 confocal microscope using a 63× water immersion objective. Images were acquired at a resolution of 1,024 × 1,024 pixels with a step size of 0.2–0.5 mm. ImageJ was used to generate maximum intensity projections of singly labelled neurons and skeletonized dendrites in the x–y and x–z planes.

Image stacks were registered using stackreg (ImageJ). Images of single VG3-ACs and W3B-RGCs taken from P30-40 retinae were skeletonized manually through the z-stack using simple neurite tracer (ImageJ). Path ROIs describing neuronal processes were converted to line stacks and used for morphological analysis using the Trees toolbox[46]. We measured the projection depth of VG3-ACs and W3B-RGCs by taking the z-axis intensity profile of labelled neurons and plotted this profile as a percentage of IPL depth. IPL depth in turn was defined by ChAT-counterstained somas that label SACs in the GCL and IPL. The investigators were not blinded to allocation during image analysis

Channelrhodopsin excitation. ChR2-tdTomato-positive interneurons were first imaged at low power (2–4 mW sample plane power at 960 nm) and a stack of their cell body positions was acquired. Somata were highlighted as regions of interest (ROIs) until all available interneurons were marked within the stack (typically 250 × 200 × 4 μm). Custom software written in LabView (National Instruments) used these ROIs to steer the beam to soma locations and activate ChR2 using either

raster or spiral scan trajectories (~25–30 mW sample plane power at 920 nm). Dwell times in these scan patterns were 0.02–0.05 ms pixel^{-1}, which was less than the rise time of the current produced by a stationary spot (Extended Data Fig. 3c, g–j). Pixel size was 0.6 μm^2 and scan patterns typically contained 100 pixels with their total time synchronized to the laser shutter, which opened for 2 ms.

We arrived at these parameters by measuring the kinetics of ChR2 responses on HEK cells and retinal neurons that expressed ChR2(H134R) (Extended Data Fig. 3). Our goal was to scan interneuron somas (typically 10×10 pixels in size) and activate them with a current step that would imitate the square current pulses used when stimulating interneurons via a sharp electrode[47]. Responses in ChR2-expressing HEK cells elicited by a stationary spot of infrared laser light at a range of wavelengths between 800 and 960 nm (840 nm, 860 nm and 920 nm are shown in Extended Data Fig. 3b, c) led us to conclude that 920 nm would be a good compromise between response size, rise time and sample plane power. Currents typically had <2 ms rise times (10–90%) with peak amplitudes of 150–175 pA at 25 mW sample plane power (Extended Data Fig. 3a). Increasing stimulus duration beyond 2–3 ms produced no increase in peak amplitude and instead the response showed signs of desensitization[48–51] (Extended Data Fig. 3a). Taken together, these results would suggest that responses evoked by 920 nm light would saturate at dwell times greater than 2 ms pixel^{-1}.

Next, we measured the spatial dimension of two-photon laser stimulation of ChR2. To do this, we recorded from ChR2 expressing HEK cells while stimulating the HEK cell at different positions away from the edge of the cell membrane (Extended Data Fig. 3d), or while stimulating the HEK cell at different heights beginning at the cell membrane (Extended Data Fig. 3f). Taken together, these results indicate that two-photon laser stimulation of ChR2 has maximal x–y–z dimensions of approximately $3 \times 3 \times 5$ μm. Next, we recorded from ChR2-positive RGCs or SACs, highlighted their somas with ROIs and steered the beam through raster or spiral scan trajectories at different dwell times to produce currents of 250–350 pA with <2ms rise times (Extended Data Fig. 3g–j). Empirically, this dwell time tended to be 0.02–0.05 ms pixel^{-1}. Currents were similar at both dwell times. Finally, we determined the relationship between inward currents on W3B to stimulation of VG3-ACs at different sample plane powers. This experiment revealed a classical sigmoid relationship between the size of the postsynaptic response and the strength of the stimulus (power at the sample plane) delivered to VG3-ACs consistent with calcium dependence of transmitter release[52] (Extended Data Fig. 3l). Sample plane powers used in our analysis of connectivity (25–30 mW) sit on the plateau of this curve and are 4–6 times the power levels that evoke a minimal response. We observed signs of infrared-light induced damage on cultured cells as well as retinal tissue at sample plane powers greater than 40 mW.

Under the conditions we used, photoreceptor-evoked light responses on RGCs were rarely seen. To confirm this independence, we activated ChR2 and photoreceptors independently and observed distinctly different kinetics[26] (Extended Data Fig. 4g). These data establish that our functional connectivity measurements are not contaminated by light responses. Nevertheless, we avoided exposure times of >100 ms owing to a well described relationship between exposure time and infrared-evoked light responses[52]. Even in these cases, however, ChR2-positive interneuron-evoked responses were clearly distinguishable from light responses owing to the marked differences in their latencies (>50 ms for light responses).

For experiments in which we activated ChR2 with blue light, we incubated the retina in a cocktail of blockers as described previously[19]. (10 mM ACET ((S)-1-(2-amino-2-carboxyethyl)-3-(2-carboxy-5-phenylthiophene-3-yl-methyl)-5-methylpyrimidine-2,4-dione)[53], Tocris; 20 mM L-AP4 (L-(+)-2-amino-4-phosphonobutyric acid), Tocris; 300 mM hexamethonium, Sigma). To activate ChR2, we focused light from a 500 mW 470 nm LED (Thor Labs) onto the retina using a relay lens and a 5× objective. LED activation was synchronized to the recordings similar to above for two-photon recordings.

Electrophysiology. Mice were dark-adapted for at least 2 h before euthanasia. The retina was rapidly dissected under infrared illumination into oxygenated (95% O$_2$; 5% CO$_2$) Ames solution (Sigma), then placed in a recording chamber with RGCs facing up. Labelled neurons were imaged under two-photon illumination and targeted for recording. For cell-attached recordings, the patch electrodes (4–7 MΩ) were filled with Ames solution. For whole-cell recordings, the electrodes were filled with an internal solution containing 120 mM K-acetate, 10 mM Na-acetate, 0.2 mM CaCl$_2$, 1 mM MgCl$_2$, 10 mM EGTA, 2 mM NaATP, 0.3 mM MgGTP, 5 mM KCl and 10 mM HEPES. Fluorescein 3,000 MW dextran (Invitrogen) was also added to the internal solution to make the electrode visible under two-photon illumination. The chloride reversal potential under these conditions is ~−70 mV, allowing good separation of inhibitory and excitatory currents in whole-cell mode. For excitatory currents, $V_h \approx -70$ mV (E_{Cl}), and for inhibitory currents, $V_h \approx -10$ mV ($E_{glutamate}$); these values were determined for each cell with a voltage ramp stimulus. Only cells with a V_m value more negative

than −50 mV and series resistances of less than 25 MΩ were used. 40 μM 2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[f] quinoxaline-7-sulfonamide (NBQX, Tocris), 100 μM picrotoxin (Tocris), 3 μM strychnine (Tocris) and CdCl$_2$ (Sigma) were dissolved in Ames. After recording, retinas were processed for histology (see above) and immunostained with antibodies against fluorescein to visualize RGC morphologies.

Signals from loose-patch and whole-cell recordings were acquired with a MultiClamp 700B amplifier (Molecular Devices) and with custom software written in LabView. For spikes, Multiclamp was put into $I = 0$ mode and signals were high pass filtered at 1 Hz. For currents, signals were digitized at 20 kHz. An unfiltered output of membrane current was taken in parallel to monitor recording quality. Series resistance for all recordings was not compensated.

Currents were processed by custom software written in Matlab (Simulink) and were analysed as follows: First, currents averaged from ~10 stimulus repetitions were split into those that evoked currents (connected) and those that did not (disconnected) in a partially automated way using the following criteria: (1) averaged traces had to have peak currents that were one standard deviation above the pre-stimulus average baseline; and (2) the variance about the mean current had to be <15% to confirm that stimulus locked currents were present in each individual trial. Once categorization was complete, traces were analysed for their rise times, decay times, onset latency, and sizes. These measurements were then remapped to the stimulated position of the interneuron. Graphs of these processed values were plotted in Igor Pro (Wavemetrics), saved as image files and inserted into figure layouts using Adobe Illustrator. A similar classification procedure was used for analysis of light evoked currents. Data from $Sdk2^{ce/+}$ and $Sdk2^{+/+}$ mice were pooled for analysis.

Visual stimuli. Visual stimuli were delivered via a projector as described[26]. In brief, all visual stimuli were written in Matlab using the psychophysics toolbox and displayed on the projector with a background intensity set to $1 \times 10^4 R^*$ rod^{-1} s^{-1}, in which R^* denotes activated rhodopsin.

The ON receptive field is larger than the OFF receptive field in W3B-RGCs[6] and slightly off-centre spots can bias the relative sizes of ON and OFF inward currents. To ensure that this innate feature of W3B-RGCs did not influence our measurements of ON and OFF current sizes, we delivered white noise stimuli to W3B-RGCs while recording inward currents. Next, we reverse correlated these currents to the stimuli and created an average that described the receptive field centre. These receptive fields were used as stimuli to measure the sizes of ON and OFF inward currents. Currents in wild-type animals had a canonical shape and showed that the peak OFF current is typically 4–8 times larger than ON current (Fig. 4).

For our differential motion stimulus, we first mapped the W3B-RGCs receptive field using a grid of flashing spots and measured spikes in $I = 0$ mode. Next, we centred a circular object region (100–150 μm) over the receptive field, bounded it with a grey annulus (50–70 μm) and created a background region that extended from the annulus to edge of our projected image. Differential motion stimuli were constructed as previously reported[6–8] and consisted of bars of 40–72 μm in width that alternated in intensity about our adapting grey level and moved at a speed of 100 μm s^{-1}, for a single bar width. For local motion, bars were passed within the object region while the background was set to grey. For global motion, bars were passed over the object and background regions. For differential motion, we passed bars over the full field but moved them in the object region at the following times before or after background movement in the background: 333.33 ms, 166.67 ms, 100 ms, 83.3 ms, 66.67 ms, 50 ms, 33.33 ms, 16.67 ms, 0 ms (Fig. 4i). These correspond to a difference of 20, 10, 5, 4, 3, 2, 1 and 0 frames of our 60-Hz projector. Responses at these different timings were collected and normalized to the response obtained under local motion only. W3B-RGCs are strongly silenced by activation of their surround, and as a result do not fire when bar movement in the two regions were synchronous. The OMS index (Fig. 4j) was computed by normalizing the average firing rate in response to differential and global motion to that measured in response to local motion. Average firing rates used for this procedure are shown in Extended Data Fig. 10i.

Diphtheria toxin injection. Diphtheria toxin (Sigma) was first dissolved in PBS at 1mg ml^{-1} and then aliquoted at −80 °C. Freshly thawed diphtheria toxin aliquot was diluted in PBS and delivered as intraperitoneal injection at 1 μg per 50 g body weight[41] to 7–10-week-old *Vglut3-cre; Rosa-LOX-STOP-LOX-DTR;TWY3* mice. The dose was repeated four times at 2-day interval. As controls, we injected diphtheria toxin in control animals and saline in *Vglut3-cre; Rosa-LOX-STOP-LOX-DTR;TWY3* animals.

AAV-mediated gene transfer. Viral-mediated gene transfer was performed as described[26]. For initial connectivity measurements, adeno-associated virus (AAV)-expressing Cre-dependent ChR2–YFP (rAAV2/9-hEF1a-DIO-ChR2 (H134R)-YFP-WPRE, AV-9-20297P, Penn Vector Core) was at a titre of ~1× 10^{13} genome copies per ml, as described previously[26]. All of these initial viral experiments were repeated using the Ai27 line. Similar results were obtained with both methods, and results from both methods were pooled.

Statistical methods. No statistical method was used to predetermine sample size. Data sets were tested for normality and statistical differences were examined using the Student's *t*-test (Igor Pro). Variance in the estimate of the mean is shown as s.e.m.

30. Chan, W. *et al.* A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic Acids Res.* **35,** e64 (2007).
31. Rodríguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genet.* **25,** 139–140 (2000).
32. Buffelli, M. *et al.* Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature* **424,** 430–434 (2003).
33. Kay, J. N. *et al.* Retinal ganglion cells with distinct directional preferences differ in molecular identity, structure, and central projections. *J. Neurosci.* **31,** 7753–7762 (2011).
34. Duan, X. *et al.* Subtype-specific regeneration of retinal ganglion cells following axotomy: effects of osteopontin and mTOR signaling. *Neuron* **85,** 1244–1256 (2015).
35. Goebbels, S. *et al.* Cre/loxP-mediated inactivation of the bHLH transcription factor gene NeuroD/BETA2. *Genesis* **42,** 247–252 (2005).
36. Kay, J. N., Voinescu, P. E., Chu, M. W. & Sanes, J. R. Neurod6 expression defines new retinal amacrine cell subtypes and regulates their fate. *Nature Neurosci.* **14,** 965–972 (2011).
37. Zhuang, X., Masson, J., Gingrich, J. A., Rayport, S. & Hen, R. Targeted gene expression in dopamine and serotonin neurons of the mouse brain. *J. Neurosci. Methods* **143,** 27–32 (2005).
38. Rossi, J. *et al.* Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell Metab.* **13,** 195–204 (2011).
39. Trenholm, S., Johnson, K., Li, X., Smith, R. G. & Awatramani, G. B. Parallel mechanisms encode direction in the retina. *Neuron* **71,** 683–694 (2011).
40. Madisen, L. *et al.* A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neurosci.* **15,** 793–802 (2012).
41. Buch, T. *et al.* A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nature Methods* **2,** 419–426 (2005).
42. Dymecki, S. M., Ray, R. S. & Kim, J. C. Mapping cell fate and function using recombinase-based intersectional strategies. *Methods Enzymol.* **477,** 183–213 (2010).
43. Furuta, Y., Lagutin, O., Hogan, B. L. & Oliver, G. C. Retina- and ventral forebrain-specific Cre recombinase activity in transgenic mice. *Genesis* **26,** 130–132 (2000).
44. Yamagata, M. & Sanes, J. R. Transgenic strategy for identifying synaptic connections in mice by fluorescence complementation (GRASP). *Front. Mol. Neurosci.* **5,** 18 (2012).
45. Lefebvre, J. L., Kostadinov, D., Chen, W. V., Maniatis, T. & Sanes, J. R. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488,** 517–521 (2012).
46. Cuntz, H., Forstner, F., Borst, A. & Häusser, M. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLOS Comput. Biol.* **6,** e1000877 (2010).
47. Asari, H. & Meister, M. The projective field of retinal bipolar cells and its modulation by visual context. *Neuron* **81,** 641–652 (2014).
48. Rickgauer, J. P. & Tank, D. W. Two-photon excitation of channelrhodopsin-2 at saturation. *Proc. Natl Acad. Sci. USA* **106,** 15025–15030 (2009).
49. Prakash, R. *et al.* Two-photon optogenetic toolbox for fast inhibition, excitation and bistable modulation. *Nature Methods* **9,** 1171–1179 (2012).
50. Andrasfalvy, B. K., Zemelman, B. V., Tang, J. & Vaziri, A. Two-photon single-cell optogenetic control of neuronal activity by sculpted light. *Proc. Natl Acad. Sci. USA* **107,** 11981–11986 (2010).
51. Katz, B. & Miledi, R. The release of acetylcholine from nerve endings by graded electric pulses. *Proc. R. Soc. Lond.* **167,** 23–38 (1967).
52. Euler, T. *et al.* Eyecup scope–optical recordings of light stimulus-evoked fluorescence signals in the retina. *Pflugers Arch.* **457,** 1393–1414 (2009).
53. Borghuis, B. G., Looger, L. L., Tomita, S. & Demb, J. B. Kainate receptors mediate signaling in both transient and sustained OFF bipolar cell pathways in mouse retina. *J. Neurosci.* **34,** 6128–6139 (2014).
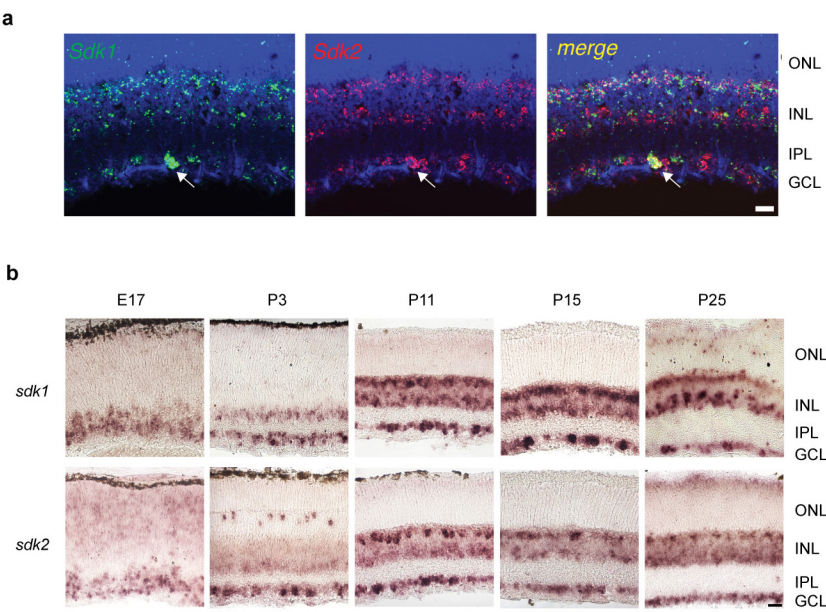
Extended Data Figure 1 | Expression of *Sdk1* and *Sdk2* in developing retina. **a**, Double-label *in situ* hybridization for *Sdk1* and *Sdk2* at P10. Arrow indicates a retinal ganglion cell that expresses both *Sdk* genes. Label in INL includes amacrine and bipolar cells, as summarized in Fig. 1k. Other images (not shown) reveal that *Sdk2* is also expressed by horizontal cells. **b**, *In situ* hybridization for *Sdk1* and *Sdk2* RNA at indicated postnatal ages. Scale bars, 10 μm.
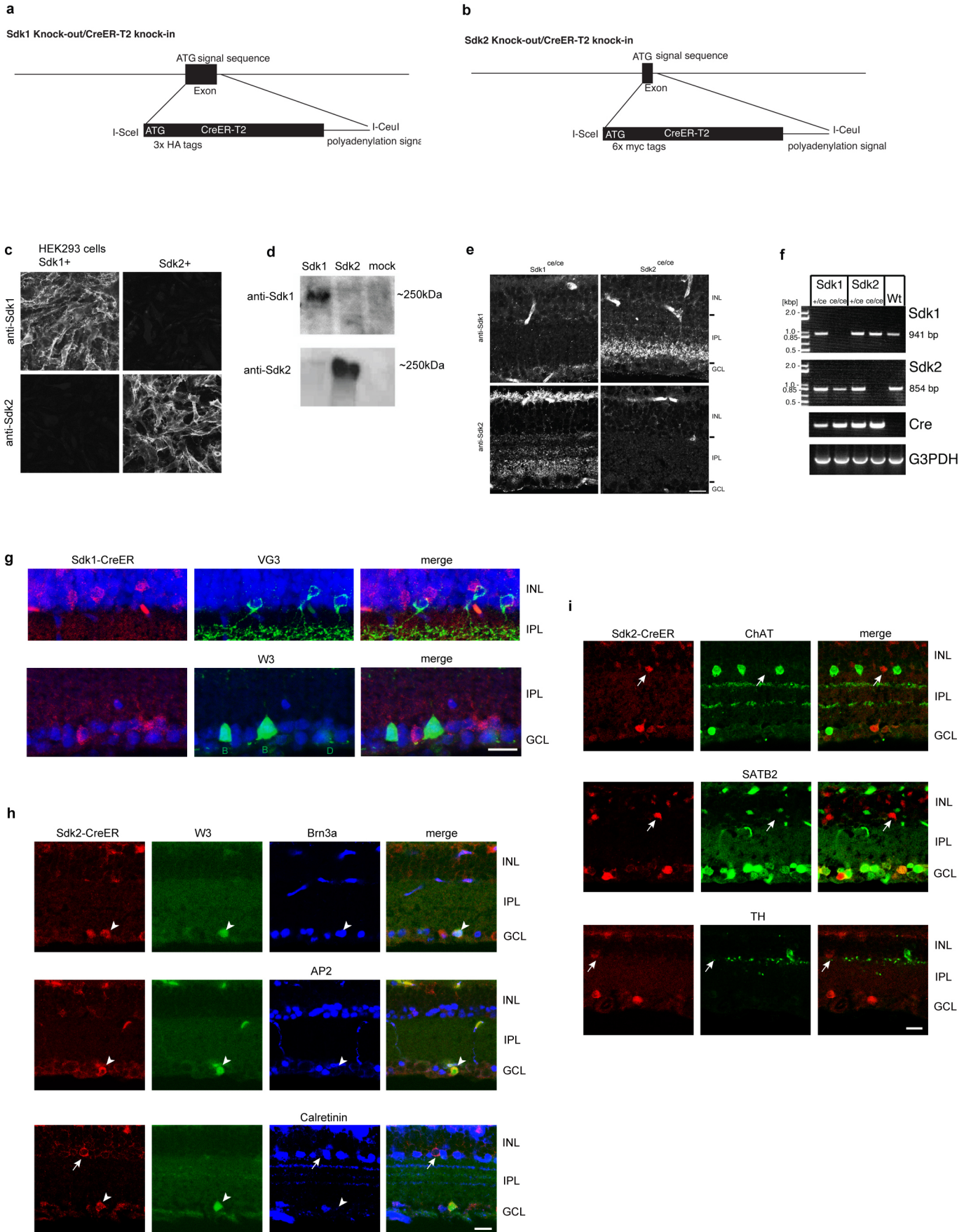
Extended Data Figure 2 | Generation and characterization of Sdk1 and Sdk2 knock-in mice a, b, Targeting vectors used to generate $Sdk1^{ce/ce}$ (**a**) and $Sdk2^{ce/ce}$ (**b**) mice. **c**, HEK293 cells transfected with expression vectors encoding $Sdk1$ or $Sdk2$, followed by staining with mouse antibodies to Sdk1 and Sdk2. **d**, HEK293 cells transfected with expression vectors encoding $Sdk1$ or $Sdk2$, followed by immunoblotting with rabbit polyclonal antibodies to Sdk1 and a mouse monoclonal antibody to Sdk2 (CS22). **e**, Retinal sections from P30 $Sdk1^{ce/ce}$ and $Sdk2^{ce/ce}$ mice stained with mouse antibodies to Sdk1 and Sdk2. Signal on blood vessels is nonspecific. **f**, Reverse transcription PCR (RT–PCR) from $Sdk1^{ce/+}$, $Sdk1^{ce/ce}$, $Sdk2^{ce/+}$ $Sdk2^{ce/ce}$ and wild-type (WT) mice. Total RNA was prepared from brain. G3PDH, glyceraldehyde-3-phosphate dehydrogenase. **g**, VG3-ACs and W3B-RGCs are Sdk1-negative. CreER expressed in P30 $Sdk1^{ce/+}$ mouse was stained with antibodies to the oestrogen receptor and vesicular glutamate transporter 3 (VG3). Bottom row shows CreER and YFP double staining in a $Sdk1^{ce/+}$; $TYW3$ mouse. **h**, Sdk2-expressing W3B cells express the RGC marker Brn3a, but not the amacrine cell marker AP2. Calretinin is expressed in all the SACs, a subpopulation of type II catecholaminergic cells (CAII) amacrine cells and some RGCs. **i**, SACs (ChAT-positive), type I catecholaminergic (tyrosine hydroxylase-positive), SEG (glycinergic) and nGnG (Satb2-positive[35]), amacrine cells do not express Sdk2. Scale bars, 10 μm.
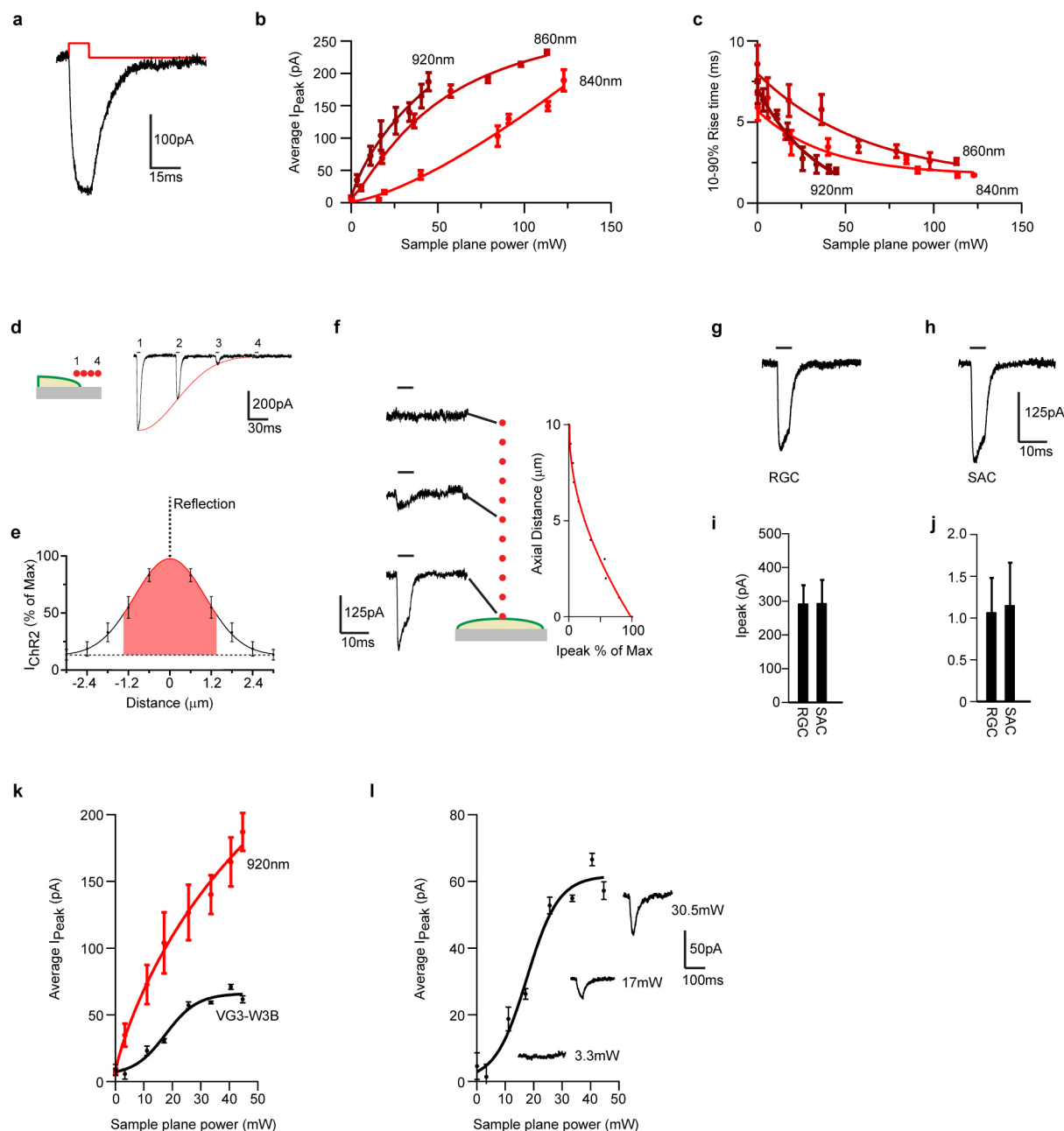
Extended Data Figure 3 | Optimization of optogenetic methods. a, Sample current recorded from a ChR2–YFP-expressing HEK cell in response to a stationary PSF-sized spot of 920-nm laser stimulation for 7 ms. **b**, Plot of average peak current measured on ChR2–YFP-expressing HEK cells in response to a stationary point spread function (PSF)-sized spot of laser stimulation at 840 nm, 860 nm and 920 nm for a range of different powers. The 920 nm light produced the largest currents for the least power. **c**, Plot of the rise time (10–90%) of currents to a stationary, PSF-sized spot of laser stimulation at 840 nm, 860 nm and 920 nm for a range of different powers. The 920 nm light produced the shortest rise time for the least power. **d**, Cartoon of a HEK cell with four adjacent PSF-sized spots that start on the edge of the cell and extend off (left). Sample currents evoked by stimulation of these regions. Currents decrease in size as the PSF moves away from the cell. **e**, Quantification of the experiment illustrated in **d** for 5 cells. The curve was obtained in only one direction and reflected about the *y* axis to give a measure of the *x*–*y* spread of ChR2 excitation. **f**, Cartoon of a HEK cell stimulated by a stationary, PSF-sized spot at 10 adjacent 1-μm planes that extend from the cell surface to −10 μm. Sample currents evoked by this procedure at the indicated *z*-positions. Average peak current measured on four cells for this procedure

(right) show the *z*-extent of ChR2 excitation. **g**, Sample current evoked in a ChR2-expressing RGC in response to stimulating an ROI (10 × 15) with a dwell time of 0.03 ms, taking a total time of 5 ms. **h**, Sample current evoked in a ChR2-expressing SAC in response to stimulating an ROI (10 × 15) with a dwell time of 0.03 ms, taking a total time of 5 ms. **i**, Average peak current evoked in RGCs and SACs in response to the stimuli shown in **g** and **h** (*n* = 6 cells). **j**, Average rise time (10–90%) of currents evoked on RGCs and SACs in response to the stimuli shown in **g** and **h** (*n* = 6 cells). **k**, Plot of average peak current measured in a W3B-RGC in response to stimulation of VG3-ACs with two-photon stimulation (920 nm) for a range of sample plane powers. Average current size shows a sigmoid relationship. Responses plateau at 25 mW sample plane power (*n* = 6 W3B-VG3 pairs in two animals). I_{peak} versus power relationship from **b** has been re-plotted for comparison. **l**, I_{peak} versus sample plane power measured on W3B-RGCs to VG3-AC replotted from **k**. Sample currents beside the curve were evoked by the powers indicated. Responses require a threshold amount of excitation in VG3-ACs; probably owing to calcium-dependent vesicle release mechanisms (*n* = 6 W3B-VG3 pairs in two animals).
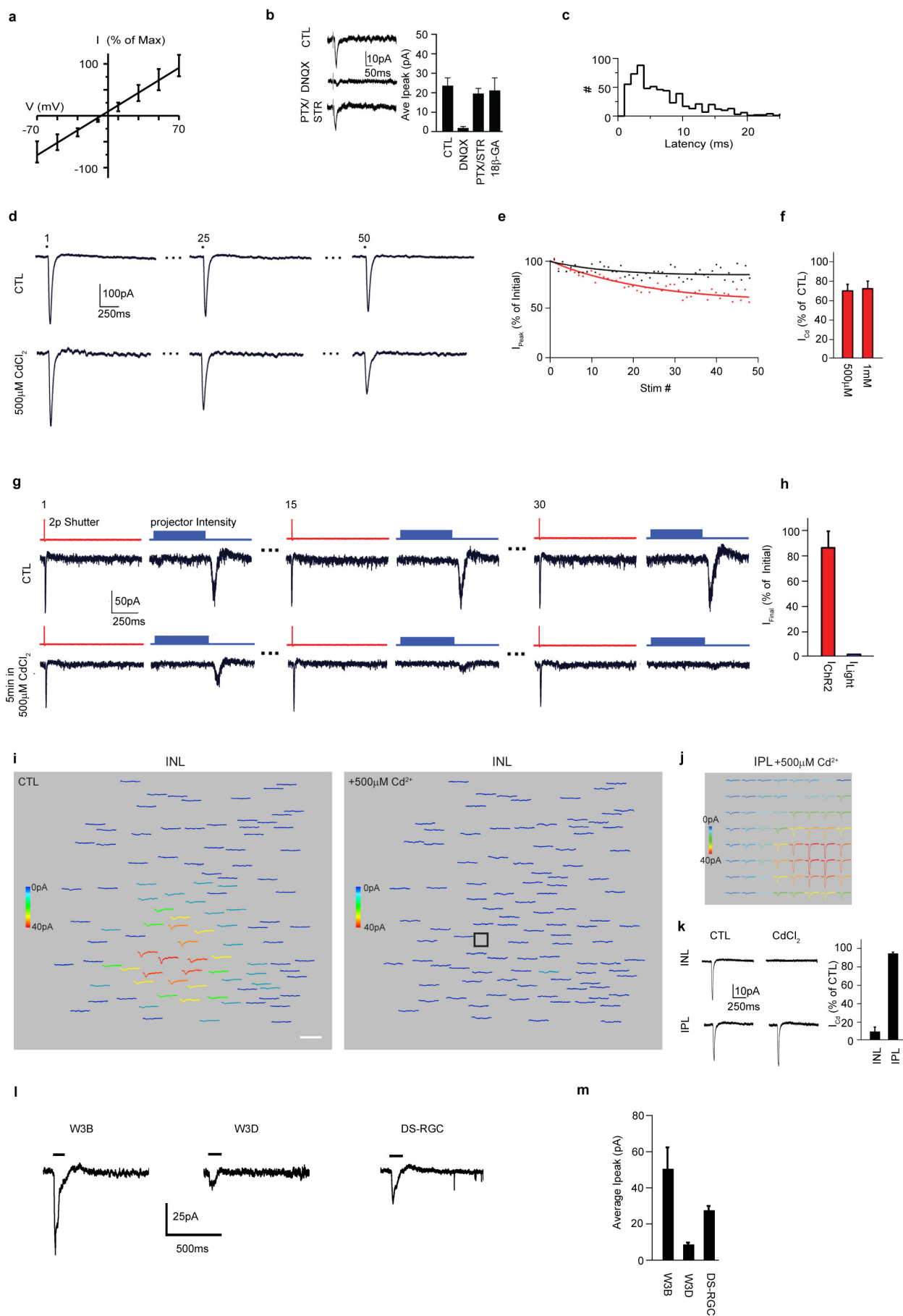
**Extended Data Figure 4 | VG3-ACs form direct excitatory synapses on W3B-RGCs. a**, Average current (I)–voltage (V) plot of VG3-AC evoked currents on W3B-RGCs. Currents were normalized to the maximum inward current per cell ($n = 10$ VG3-W3B pairs). **b**, Currents are abolished by inhibitors of AMPA-type glutamate receptors (DNQX, 20 μM), but are unaffected by inhibitors of GABA (picrotoxin (PTX), 100 μM), glycine (strychnine (STR), 3 μM) receptors and gap junctions (18β-glycyrrhetinic acid (18β-GA), 25 μM) ($n = 8$–10 per condition). **c**, Latency of VG3-AC-evoked currents on W3B ($n = 623$ VG3s and 14 RGCs). **d**, Inward currents measured in HEK cells transiently transfected with constructs containing ChR2–GFP. The first, twenty-fifth and fiftieth responses to a train (1 Hz) of 60 stimuli are shown for HEK cells recorded in control and CdCl$_2$-containing solution. **e**, Peak currents from the experiment shown in **a**, plotted as a percentage of the initial peak size. Currents in CdCl$_2$ appear to decrease slightly over time. **f**, Quantification of peak currents measured in Cd^{2+}-containing solution expressed as a percentage of those found in control solution. ChR2 activity is largely unaffected by the presence of CdCl$_2$ ($n = 6$ cells, in control solution, 10 cells in 500 μM and 8 cells in 1 mM CdCl$_2$ solution) **g**, Responses of a ChR2-positive RGC to stimulus train that alternated between two-photon excitation of ChR2 on the RGC soma and one-photon activation of photoreceptors above the RGC. The first, fifteenth and thirtieth responses in the train are shown. ChR2 responses are insensitive to the calcium channel blocker but light responses are not. **h**, Peak current measured on the final pulse of the train expressed as a percent of the first for the experiments performed like that in **e**. I_{ChR2} is largely immune to the calcium channel blocker CdCl$_2$ ($n = 4$ ChR2-positive RGCs). **i**, Heat map of VG3-AC responses measured in a W3B-RGC in control solution (left) and the same heat map measured in the presence of 500 μM CdCl$_2$. VG3-AC responses initiated by soma stimulation require functional voltage gated calcium channels in the nerve terminals. **j**, Heat map of VG3-AC responses measured in the same W3B-RGC in CdCl$_2$ solution shown

in **f** in response to a 10 × 10 stimulus grid (grid square = 5 × 5 pixels) in the inner plexiform layer. Activating ChR2 on VG3-AC nerve terminals produces responses in W3B-RGCs in spite of silencing voltage-activated calcium channels globally. Scale bars, 65 ms and 34 μm. **k**, Currents recorded from W3B-RGCs (−60 mV) in response to stimulation of VG3-ACs at their somas (INL) or terminals (IPL) in control and CdCl$_2$ (200μM) containing solution (left). Average peak current evoked by either soma (INL) or nerve terminal (IPL) stimulation in CdCl$_2$ solution expressed as a percentage of that found in control (right, $n = 276$ VG3-ACs and 5 W3B-RGCs). Currents evoked by IPL stimulation in CdCl$_2$-containing solution result from Ca^{2+} influx via ChR2 on VG3-AC terminals. VG3-ACs synapse directly with W3B-RGCs. **l**, Sample currents evoked in W3B-RGCs, W3D-RGCs and ooDSGCs by blue light (one-photon) stimulation of ChR2 positive VG3-ACs. Experiments were done in the presence of a cocktail of blockers: 10 μM ACET, to block the OFF pathway; 20 μM L-AP4, to block the ON pathway, and 300 μM hexamethonium, to block cholinergic nicotinic receptors[19,51]. **m**, Average peak currents from experiments like those in **l** ($n = 15$ W3B-RGCs, $n = 7$ W3D-RGCs and $n = 13$ ooDSGCs). These results from one-photon stimulation of a population of VG3-ACs, confirm the conclusion from two-photon stimulation of single VG3-ACs (Fig. 2): these amacrine cells innervate W3B-RGCs far more strongly than W3D-RGCs or ooDSGCs. By one-photon stimulation, the currents evoked in ooDSGCs are stronger than those in W3D-RGCs, whereas they are similar in the two-photon data. This difference probably arises from their larger dendritic size. On the basis of the dendritic diameter of VG3-ACs (~50 μm; Figs 2c and 3r), W3B-RGCs (~115 μm; Figs 2c and 3l), W3D-RGCs (~125 μm; data not shown) and ooDSGCs (200 μm; ref. 37), we can estimate that dendrites of W3B-RGCs, W3D-RGCs and ooDSGCs overlap dendrites of ~24, ~29 and ~100 VG3-ACs, respectively. Given the percentage connectivity in shown in Fig. 2, we estimate that each W3D-RGC is innervated by 10–15 VG3-ACs, whereas each ooDSGC is innervated by ~35 VG3-ACs.
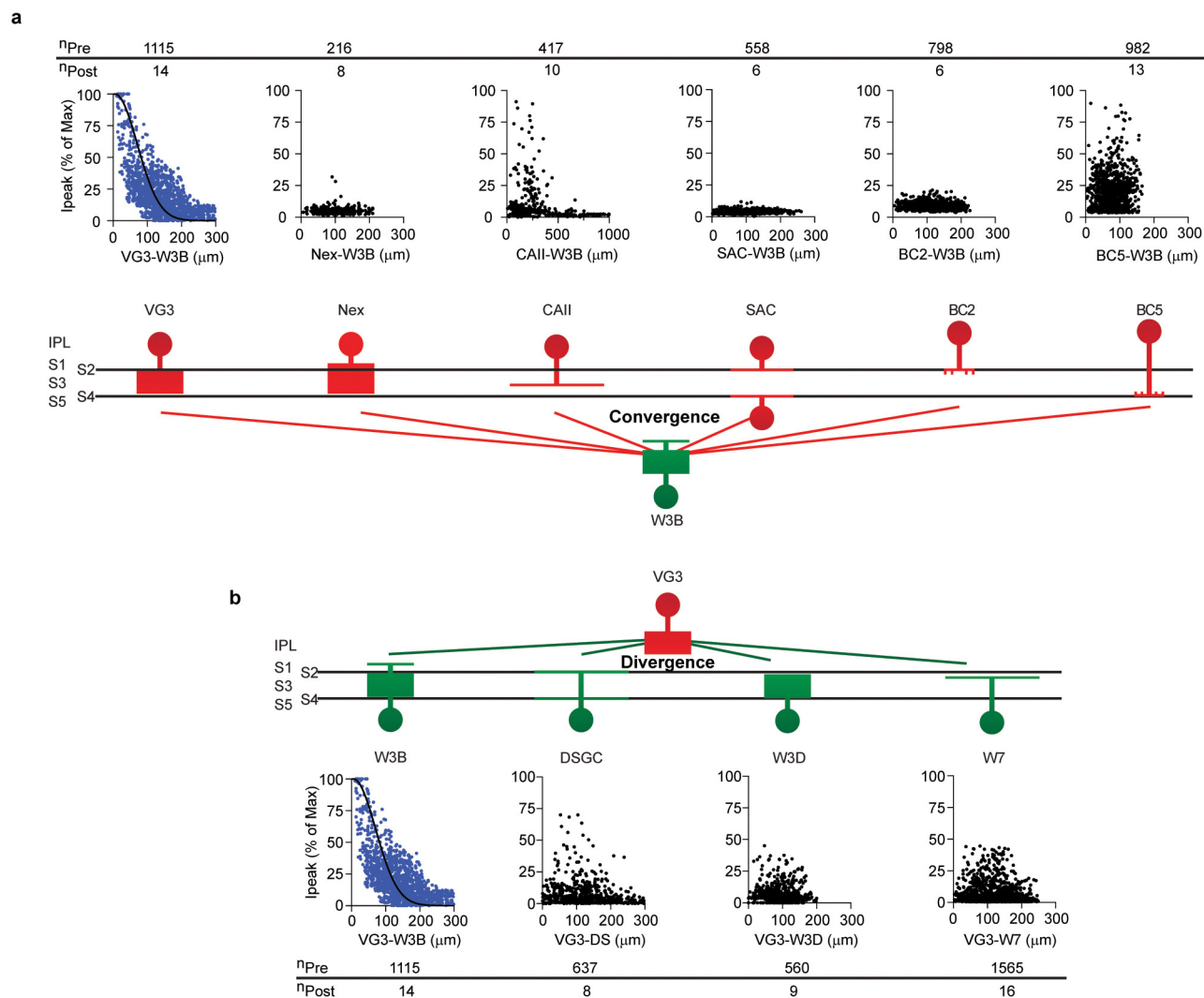
Extended Data Figure 5 | Synaptic connectivity of VG3-ACs and W3B-RGCs. **a**, **b**, Strength of connections as a function of distance from six interneuron types to W3B-RGCs (**a**) and VG3-ACs to four RGC types (**b**). Number of synaptic partners assayed shown above and sample currents shown below each graph. The W7 population contained six nearly-disconnected and ten connected pairs, presumably corresponding to the S1-laminating and S3-laminating W7 subsets. Normalized peak currents (I_{peak}) from each pair were normalized to the average maximum response from VG3-AC–W3B-RGCs.
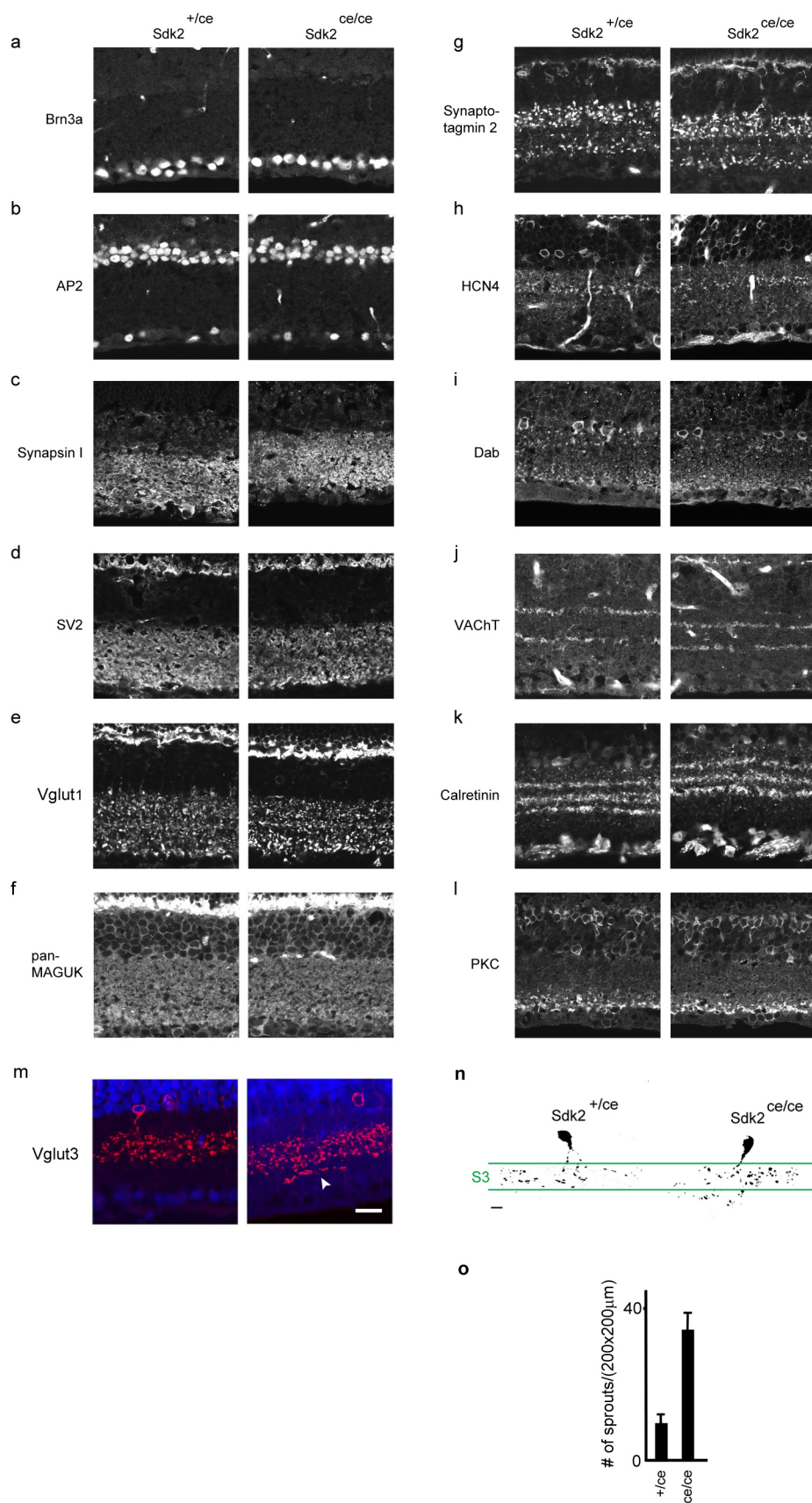
**Extended Data Figure 6 | Normal retinal architecture in Sdk2 mutants.**
Sections of $Sdk2^{ce/+}$ and $Sdk2^{ce/ce}$ retinas (P30) were stained with antibodies to cell-type-specific markers, synaptic components or fluorescent proteins. No differences between mutants and controls were detected in cells other than VG3-ACs and W3B-RGCs. **a**, Brn3a labels most RGCs. **b**, AP2 labels all amacrine cells. **c, d**, Synapsin I and SV2 are associated with synaptic vesicles. **e**, VGlut1 is concentrated in photoreceptor and bipolar terminals. **f**, PSD-95 family members, stained with anti-MAGUK, are associated with synaptic sites, usually excitatory postsynaptic densities. **g**, Synaptotagmin 2 is concentrated in bipolar cells types 2 and 6. **h**, Anti-HCN4 labels type 3a bipolar cells. **i**, Anti-disabled-1 (Dab1) labels AII amacrines. **j**, Anti-VAChT labels dendrites of SACs. **k**, Anti-calretinin labels subsets of RGCs and amacrine cells, including SACs. **l**, Anti-protein kinase C-α (PKCα) labels rod bipolar cells and a subset of amacrine cells. **m**, Anti-VGlut3 labels VG3-ACs. Sprouting is evident in the mutant. **n**, Single-cell reconstructions of VG3-ACs labelled sparsely with GFP in retinal cross-sections. **o**, Quantification of sprouting in mutant VG3-ACs. Scale bars, 10 μm.
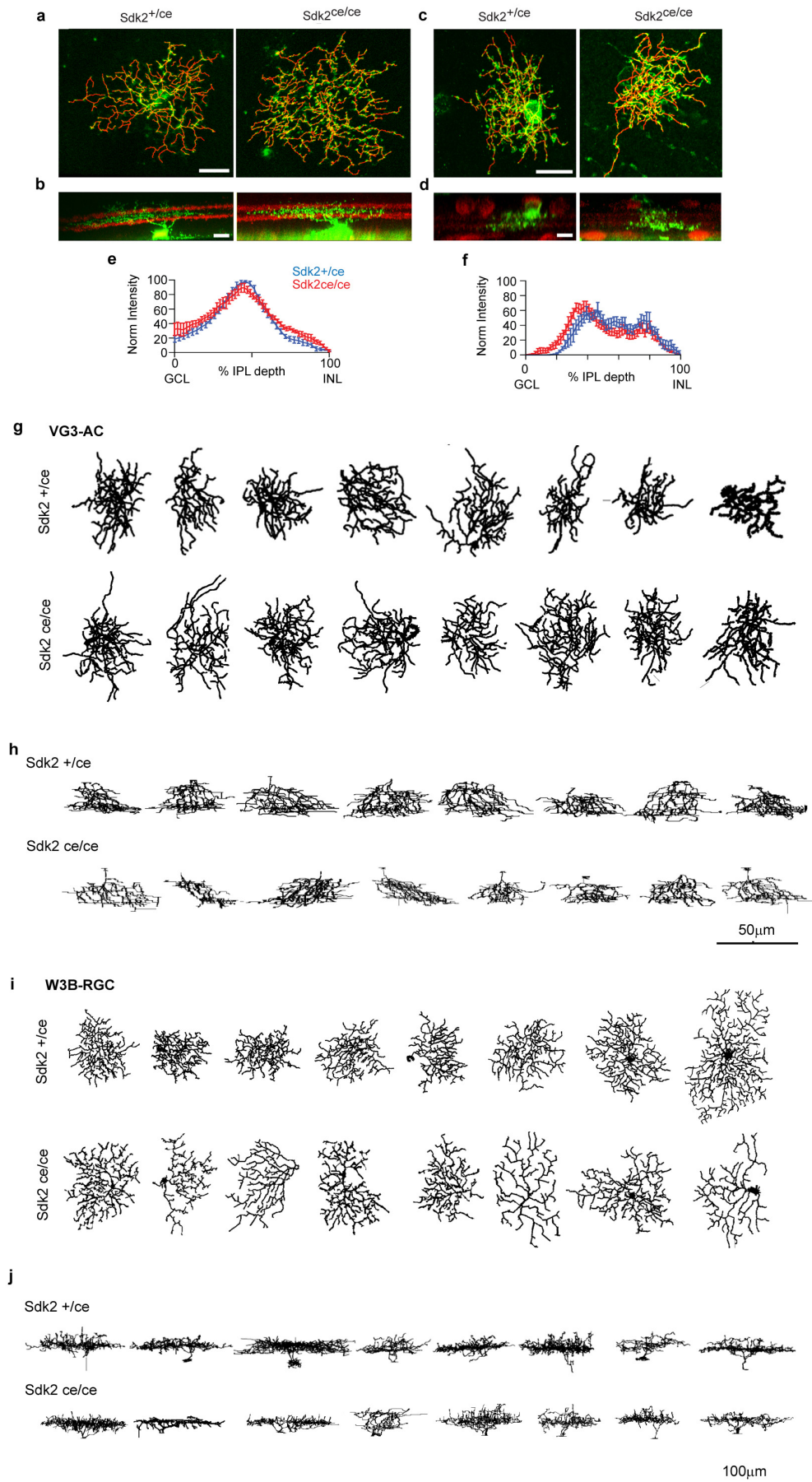
Extended Data Figure 7 | Morphological analysis of VG3-ACs and W3B-RGCs in *Sdk2* mutants. a, b, Dye-injected W3B-RGCs were imaged and skeletonized as described in Methods. Projections on rotated stacks counterstained with anti-VAChT antibodies are shown in **h. c, d,** Similar to **a, b,** except for VG3-ACs labelled sparsely with tdTomato and counterstained with anti-ChAT to label the somas of SACs. Scale bars, 25 μm. **e, f,** Mean intensity ($\pm$s.e.m.) of dye-labelled W3B-RGC dendrites (**e**) and VG3-AC dendrites (**f**) across the IPL from images such as those shown in **a–d. g, h,** En face or laminar projections of skeletonized dendritic arbors from VG3-ACs labelled sparsely with tdTomato in $Sdk2^{+/ce}$ and $Sdk2^{ce/ce}$ retinae. **i, j,** En face or laminar projections of skeletonized dendritic arbors from dye-filled W3B-RGCs in $Sdk2^{+/ce}$ and $Sdk2^{ce/ce}$ retinae.
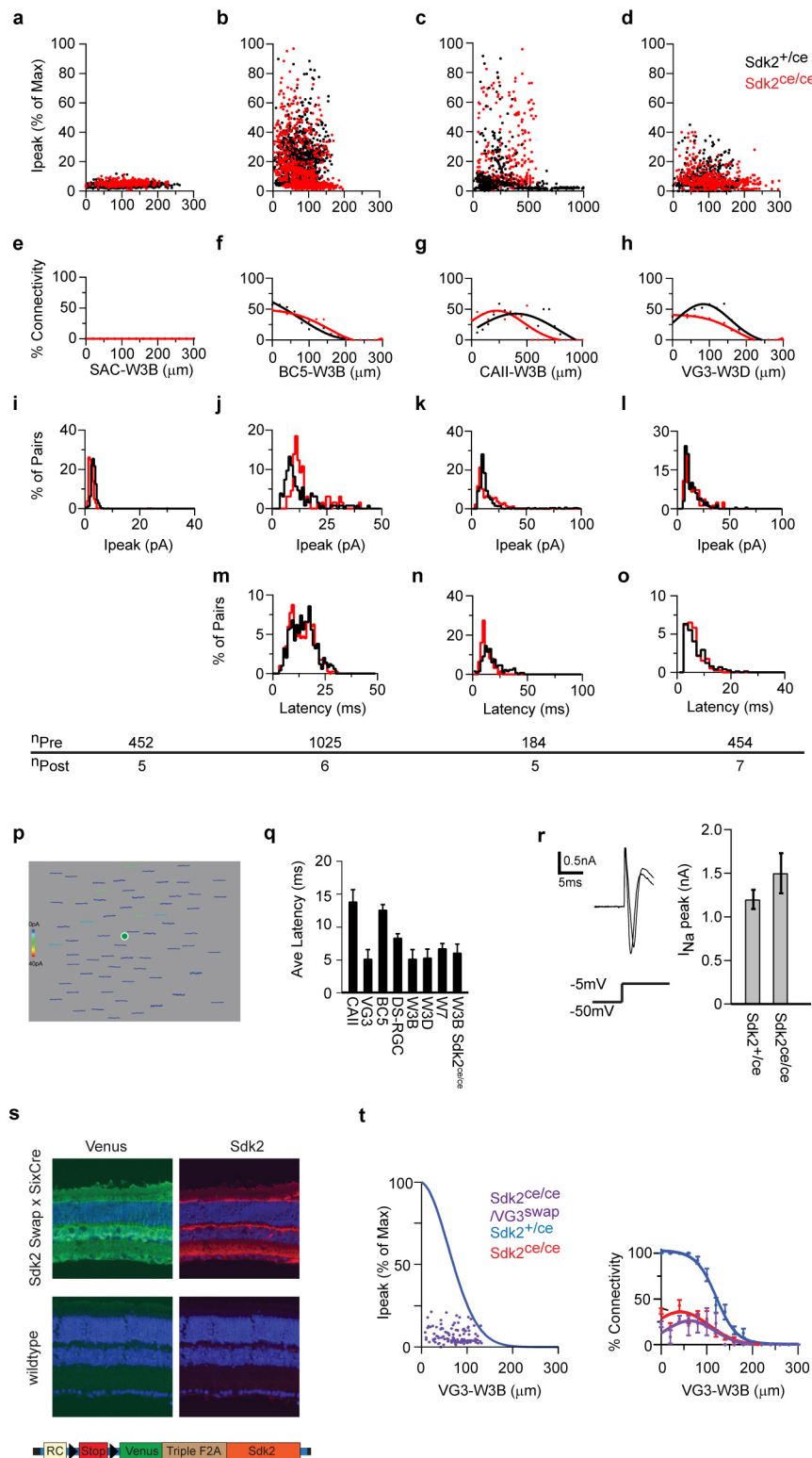
Extended Data Figure 8 | Electrical and synaptic properties of *Sdk2* mutant VG3-ACs and W3B-RGCs. a–l, Strength of connections as a function of distance from three interneuron types (SACs, type 5 bipolar cells (BC5) and type II catecholaminergic cells (CAII)) to W3B-RGCs and from VG3-ACs to W3D-RGCs in *Sdk2*$^{ce/+}$ (black) and *Sdk2*$^{ce/ce}$ (red) retinae. **a–d**, Normalized peak current. **e–h**, Percentage of connected pairs. **i–l**, Average current amplitudes in a 30-ms window after the stimulus pulse. **m–o**, Latencies of currents detectable above noise. Number of pre- and postsynaptic partners assayed shown below each column. **p**, Responses evoked in a W3B-RGC (green dot) in a *Sdk2* mutant after stimulation of 66 VG3-ACs (genotype: *TYW3; Sdk2*$^{ce/ce}$; *Vglut3-cre; LOX-STOP-LOX-ChR2-tdTomato*). **q**, Average latencies

for currents detectable above noise in control. **r**, Sodium currents in W3B-RGCs *Sdk2*$^{ce/+}$ and *Sdk2*$^{ce/ce}$ retinae. Sample currents evoked by a step from −60 mV to −5 mV (left), and average peak sodium current amplitude measured on W3B-RGCs in *Sdk2*$^{ce/+}$ and *Sdk2*$^{ce/ce}$ retinae. **s**, Retinal cross-sections from wild-type mice and those that overexpress *Sdk2* broadly using the *Six3-cre* driver and the *Sdk2* swap transgene. *Sdk2* is expressed strongly from the swap transgene, which is shown as a schematic below the micrographs. **t**, Strength of connections from VG3-ACs to W3B-RGCs in *Sdk2*$^{ce/ce}$ mutants in which Sdk2 expression was rescued in VG3-ACs. Data were derived from 121 VG3-ACs and 3 W3B-RGCs. Fit to the control data (Fig. 3a) re-plotted in blue.
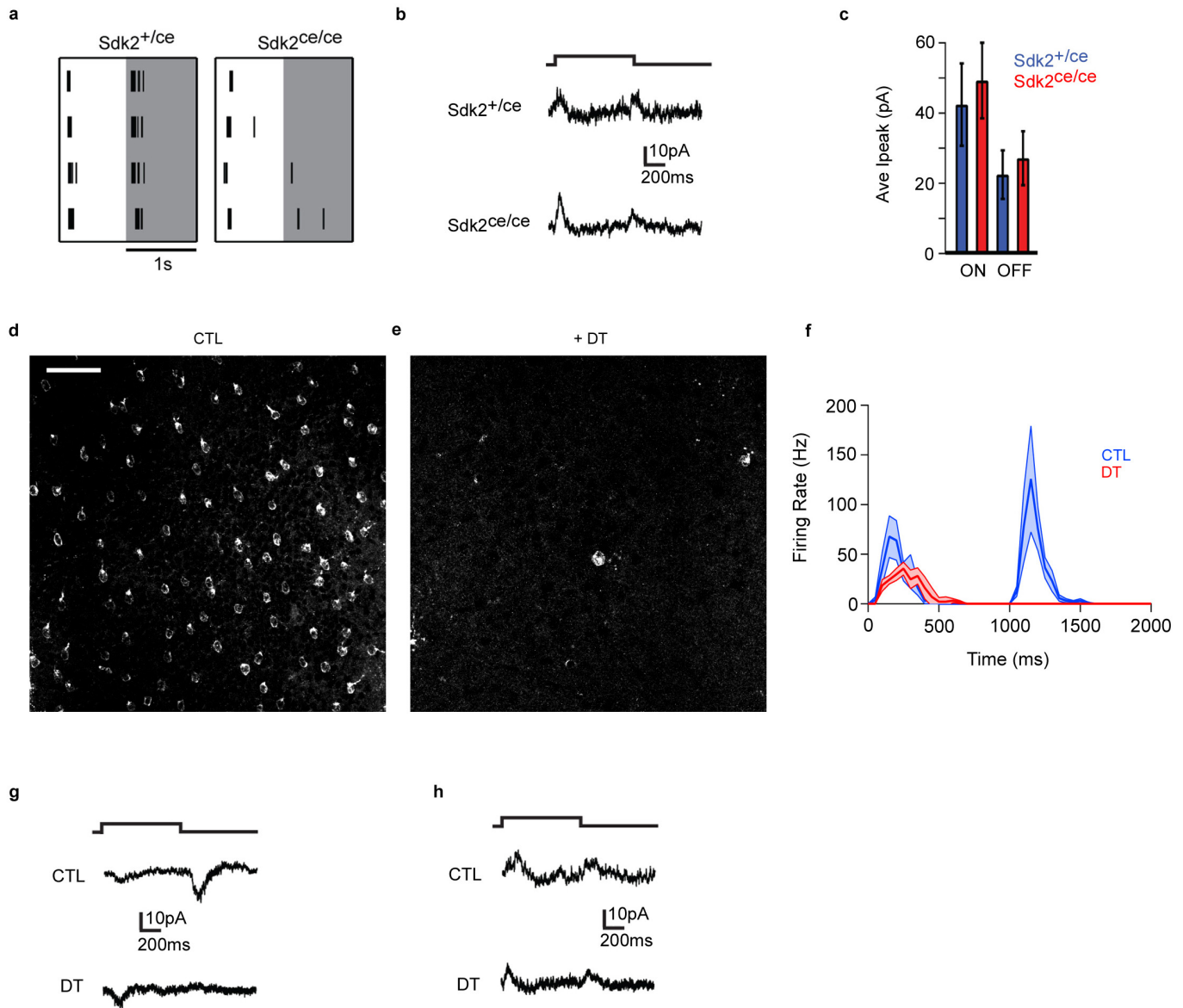
Extended Data Figure 9 | Effective deletion of VG3-ACs by diphtheria toxin. a, Spike responses of W3B-RGCs in $Sdk2^{ce/+}$ and $Sdk2^{ce/ce}$ retinae in response to a ~100-μm flashing spot centred on the receptive field. OFF responses are strongly reduced in the absence of Sdk2. **b**, Inhibitory currents recorded ($V_h = -5$ mV) from W3B-RGCs in $Sdk2^{+/ce}$ and $Sdk2^{ce/ce}$ retinae in response to a full field flash for 1 s. **c**, I_{peak} measured from excitatory currents in non-W3B-RGCs ($n = 8$ in 4 $Sdk2^{ce/+}$ mice and 8 in 4 $Sdk2^{ce/ce}$ mice). **d, e**, Sample images of tdTomato-positive VG3-ACs in retinae from diphtheria-toxin-treated control (**d**) and $Vglut3$-$cre; Rosa$-CAG-LOX-$STOP$-LOX-DTR

mice (**e**). Scale bar, 40 μm. **f**, Average firing rate recorded from W3B-RGCs in control (blue, $n = 21$) and diphtheria-toxin-treated $Vglut3$-$cre;DTR/TYW3$ (red, $n = 12$) retinae in response to small spots flashed for 1 s. Dark lines, average; shadowing denotes s.e.m. Bin width, 50 ms. **g**, Excitatory currents recorded ($V_h = -65$ mV) from W3B-RGCs in control and diphtheria-toxin-treated $Vglut3$-$cre;DTR/TYW3$ retinae in response to small spots flashed for 1 s. **h**, Inhibitory currents recorded ($V_h = -5$ mV) from W3B-RGCs in control and diphtheria-toxin-treated $Vglut3$-$cre;DTR/TYW3$ retinae in response to small spots flashed for 1 s.
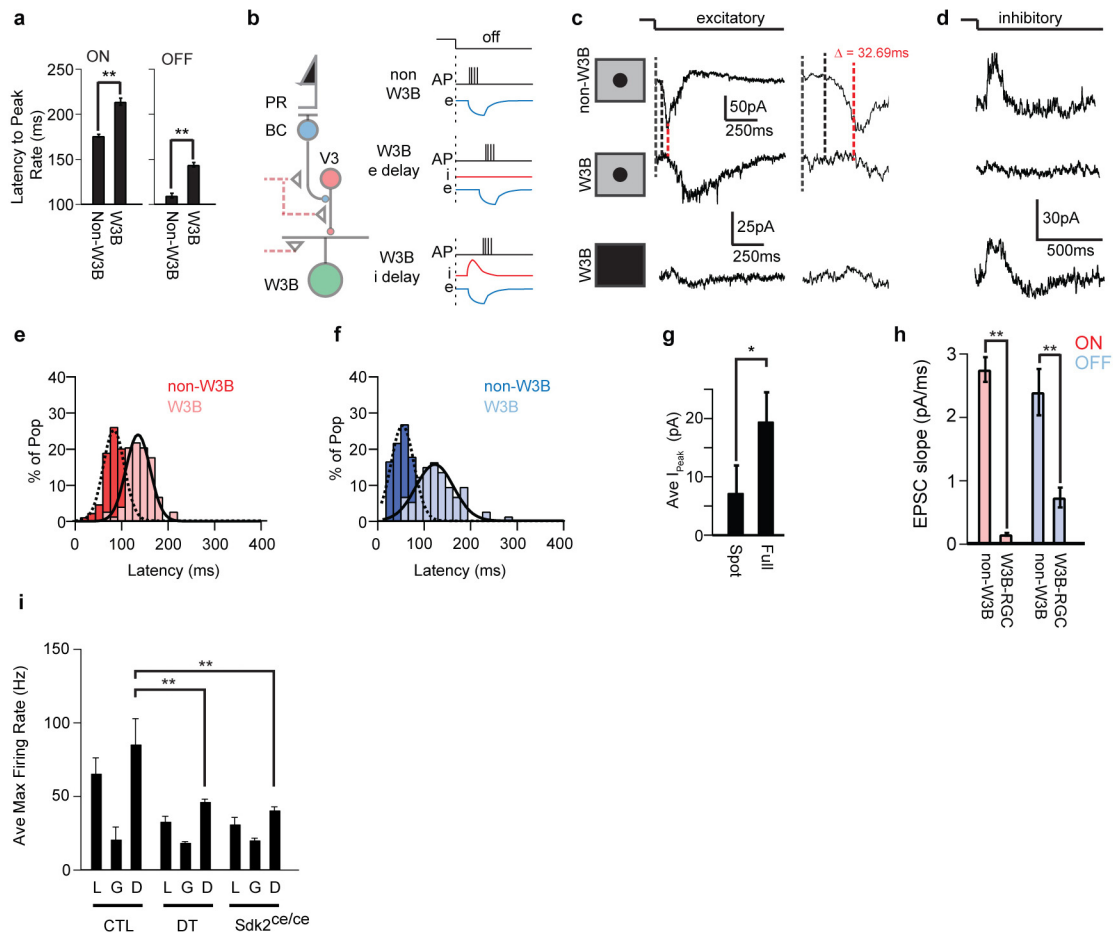
Extended Data Figure 10 | A delay line in the differential motion response.
a, Average latency to peak firing rate in W3B-RGCs and non-W3B-RGCs in response to a spot flashed over their receptive field centre ($n = 21$ non-W3B-RGCs and 18 W3B-RGCs). Spike responses on W3B-RGCs are delayed (**$P < 0.01$, Student's t-test). **b,** Possible mechanisms for the delay. In non-W3B-RGCs excitatory currents produced by bipolar cells drive the neuron to fire. In W3B-RGCs, spikes could be delayed because the interposition of VG3-ACs delays the onset of the excitatory postsynaptic current (EPSC) (W3B e delay) or because of a transient inhibitory postsynaptic current that arrives at the same time as the EPSC and delays the cell from reaching threshold (W3B i delay). **c, d,** Excitatory (**c**) and inhibitory (**d**) currents measured from non-W3B-RGCs and W3B-RGCs in response to a flashing spot centred on the receptive field or a full field flash. Dotted lines denote the stimulus onset, non-W3B-RGC EPSC onset and W3B-RGC EPSC onset (red). The onset of EPSCs in W3B-RGC lags behind those found on non-W3B-RGCs by ∼30 ms. No significant transient inhibition was observed in the receptive field centre (**d**). **e, f,** Histogram of latency to the onset of the ON (**e**) and OFF (**f**) excitatory current measured on W3B-RGCs (light blue, light red) and non-W3B-RGCs (blue, red). W3B-RGCs lag non-W3B-RGCs by ∼32–40 ms and also have higher variance about the mean ($n = 29$ non-W3B-RGCs and 27 W3B-RGCs). **g,** Average peak of inhibitory currents measured on W3B-RGCs in response to a ∼100-μm diameter flashing spot and a full field flash ($n = 6$ wild-type W3B-RGCs; *$P < 0.05$, Student's t-test). **h,** Slope of the light-evoked excitatory current between 10% and 90% of the peak. Slopes of currents in W3B-RGCs are significantly lower than those of non-W3B-RGCs ($n = 29$ non-W3B-RGCs and 27 W3B-RGCs; **$P < 0.001$). The non-W3B-RGCs included W3D-RGCs, ooDSGCs, α-RGCs and some unidentified RGCs. **i,** Average maximal firing rates measured in W3B-RGCs to local, global and differential motion stimuli ($n = 10$ W3B-RGCs in 5 control mice, 6 W3B-RGCs in 4 $Sdk2^{ce/ce}$ mice and 9 W3B-RGCs in 3 diphtheria-toxin-treated $Vglut3$-$cre;DTR/TYW3$ mice; **$P < 0.01$, Student's t-test). Firing rates in global and differential motion were normalized to that elicited in control W3B-RGCs with local motion stimuli for Fig. 4j.

LETTER

SEC14L2 enables pan–genotype HCV replication in cell culture

Mohsan Saeed[1], Ursula Andreo[1], Hyo-Young Chung[1], Christine Espiritu[1], Andrea D. Branch[2], Jose M. Silva[3] & Charles M. Rice[1]

Since its discovery in 1989, efforts to grow clinical isolates of the hepatitis C virus (HCV) in cell culture have met with limited success. Only the JFH-1 isolate has the capacity to replicate efficiently in cultured hepatoma cells without cell culture-adaptive mutations[1–3]. We hypothesized that cultured cells lack one or more factors required for the replication of clinical isolates. To identify the missing factors, we transduced Huh-7.5 human hepatoma cells with a pooled lentivirus-based human complementary DNA (cDNA) library, transfected the cells with HCV subgenomic replicons lacking adaptive mutations, and selected for stable replicon colonies. This led to the identification of a single cDNA, *SEC14L2*, that enabled RNA replication of diverse HCV genotypes in several hepatoma cell lines. This effect was dose-dependent, and required the continuous presence of *SEC14L2*. Full-length HCV genomes also replicated and produced low levels of infectious virus. Remarkably, SEC14L2-expressing Huh-7.5 cells also supported HCV replication following inoculation with patient sera.

Mechanistic studies suggest that SEC14L2 promotes HCV infection by enhancing vitamin E-mediated protection against lipid peroxidation. This provides a foundation for development of *in vitro* replication systems for all HCV isolates, creating a useful platform to dissect the mechanisms by which cell culture-adaptive mutations act.

Hepatitis C virus (HCV) is a leading cause of liver disease worldwide, with an estimated global burden of 185 million chronic infections[4]. Studies of this important human pathogen have historically been hampered by the lack of a cell culture system that supports replication of clinical isolates. Inoculation of primary human hepatocytes or hepatoma cell lines with serum from HCV-infected patients leads to extremely little, if any, replication. In addition, molecularly cloned HCV genomes that are infectious in experimentally infected chimpanzees fail to establish infection in Huh-7-derived human hepatoma cells[5–8]. Drug-selectable subgenomic replicons derived from these genomes must acquire mutations to replicate in cell culture[1,2].
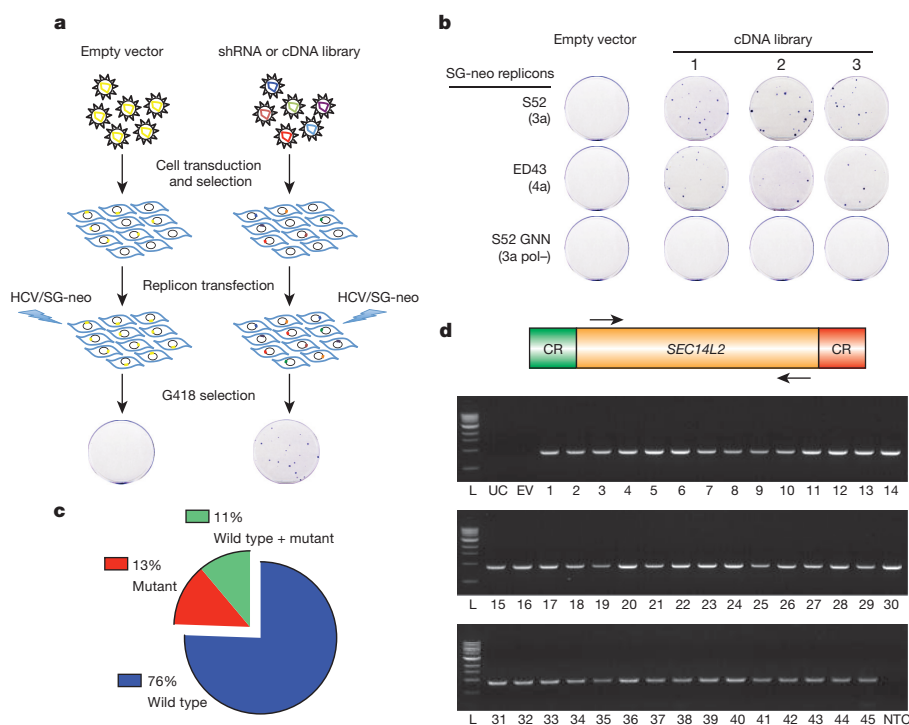

Figure 1 | cDNA screening of Huh-7.5 cells identifies SEC14L2 as a critical host factor for HCV RNA replication. a, Illustration of the shRNA or cDNA screen. **b**, Three Huh-7.5 cell populations independently transduced with a pooled lentiviral cDNA library and a population of empty-vector-expressing control cells was electroporated with the indicated wild-type HCV subgenomic replicons (SG-neo) or a replication-defective S52 GNN replicon (3a pol−). After 3 weeks of G418 selection, the resulting cell colonies were stained with crystal violet. **c**, HCV NS3-NS5B region from 45 cell colonies was amplified by reverse transcription followed by PCR (RT–PCR) and subjected to direct sequencing. Shown are the per cent colonies harbouring wild-type, mutant, or a mixture of wild-type and mutant sequences. **d**, The *SEC14L2* cDNA was amplified from 45 colonies (see Methods). CR, constant region; L, ladder; UC, untransduced cells; EV, empty-vector-transduced cells; NTC, no template control.

[1]Center for the Study of Hepatitis C, Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, New York 10065, USA. [2]Division of Liver Diseases, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [3]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

This replication block could be due to the presence of inhibitory factor(s) that limit replication and/or the lack of essential host factor(s).

To overcome the block, we transduced Huh-7.5 cells with lentivirus libraries expressing either short hairpin RNAs (shRNAs) or complementary DNAs. The transduced cell populations were then electroporated with *in vitro* transcripts of wild-type G418-selectable genotype 3a and 4a HCV subgenomic replicons, which in unaltered Huh-7.5 cells require at least two adaptive mutations for replication (Fig. 1a). While the shRNA library did not yield any hits, the cDNA library produced numerous colonies (Fig. 1b). The vast majority of these colonies (34 of 45) harboured replicons with the parental sequence (Fig. 1c); mutations were present in the remaining colonies but none corresponded to known adaptive changes. To confirm their ability to support non-adapted replicons, cell colonies were treated with anti-HCV compounds to clear replicating viral genomes and then re-transfected with a second set of wild-type replicons. Selection with G418 resulted in the production of numerous G418-resistant colonies. In contrast, no colonies were produced from control cells cured of cell culture-adapted replicons (Extended Data Fig. 1).

PCR amplification and sequence analysis of integrated cDNAs from 45 colonies revealed the same gene product, SEC14L2 (Fig. 1d). SEC14L2, also known as c22orf6, supernatant protein factor 1 (SPF1), or tocopherol-associated protein 1 (TAP1), is a cytosolic lipid-binding protein family member[9,10], and is ubiquitously expressed in human tissues[11]. *SEC14L2* RNA and protein could not be detected in human hepatoma and non-hepatoma cell lines. However, primary human hepatocytes, both from fetal and adult sources, expressed readily detectable levels (Extended Data Fig. 2a, b).

To confirm that SEC14L2 is necessary and sufficient for HCV RNA replication, we generated Huh-7.5 cells stably expressing SEC14L2 (SEC14L2/Huh-7.5) and transfected them with a panel of wild-type replicons from HCV genotypes 1a, 1b, 2a, 3a, 4a, and 5a. Selection with

G418 yielded large numbers of colonies (Fig. 2a), with the majority harbouring replicons with the parental sequences (Extended Data Fig. 2c). HCV RNA levels in these colonies were comparable to cell culture-adapted replicons, suggesting high levels of replication (Fig. 2b). The effect was not cell-line-specific, since SEC14L2 expression in Huh-7 and Hep3B/miR122 cells also rendered them permissive for HCV replication, albeit to lower levels (Extended Data Fig. 2d).

To examine the temporal and dose-dependent relationship between SEC14L2 expression and HCV permissiveness in Huh-7.5 cells, we established a doxycycline-inducible expression system. Two cell clones with varying levels of SEC14L2 expression were tested for replicon colony formation. The number of G418-resistant colonies correlated with the level of SEC14L2, indicating a dose-dependent effect (Fig. 2c and Extended Data Fig. 3a, b). Withdrawal of doxycycline led to the loss of colonies, suggesting that continuous expression of SEC14L2 is needed to maintain HCV replication (Fig. 2d).

In contrast to wild-type HCV replicons, cell culture-adapted replicons showed a mild to moderate response to SEC14L2 expression (Extended Data Fig. 3c), indicating that cell culture-adaptive mutations partially circumvent the need for SEC14L2. This notion is further supported by the fact that wild-type replicons in the presence of SEC14L2 yielded almost as many colonies as were obtained for cell culture-adapted replicons in the absence of SEC14L2. Similarly, replication of the JFH-1 subgenomic replicon, which does not require adaptive mutations, was not enhanced by SEC14L2 (data not shown).

To examine replication of full-length HCV genomes, we took two approaches. First, we performed transient replication assays in Huh-7.5 cells stably expressing an HCV infection-dependent fluorescence relocalization (HDFR) reporter[12], with or without *SEC14L2*. HCV RNA replication in HDFR cells results in nuclear localization of otherwise mitochondrially anchored red fluorescent protein (RFP). When *in vitro* transcribed full-length HCV RNA genomes were
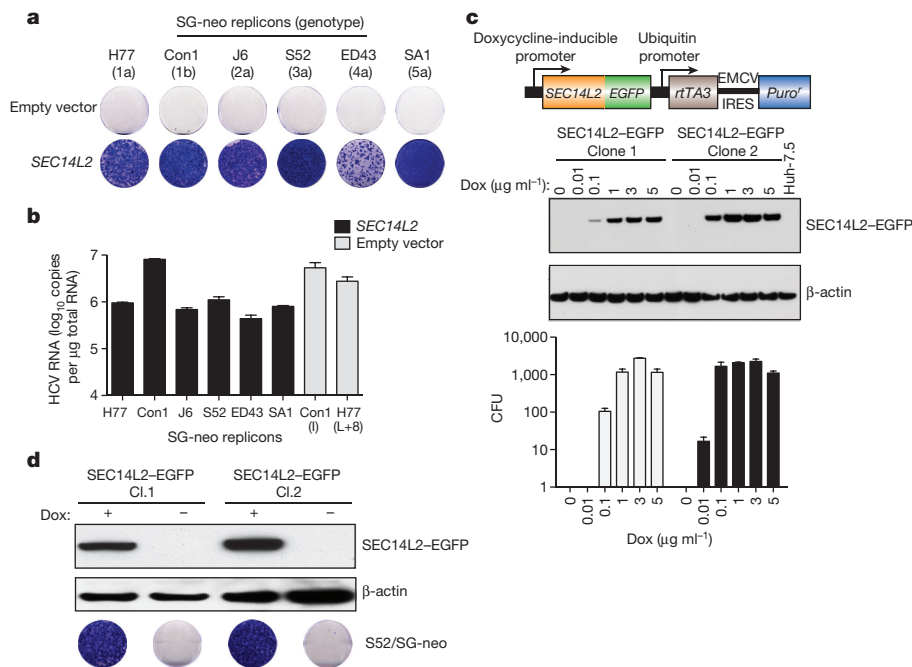

Figure 2 | *SEC14L2* allows replication of wild-type HCV subgenomic replicons. a, Cell colonies obtained for the indicated replicons in Huh-7.5 cells transduced with empty vector or *SEC14L2*. **b**, HCV RNA levels for wild-type replicons in SEC14L2-expressing cells (black bars) and two cell culture-adapted replicons in control cells (grey bars) are shown. **c**, Schematic representation of the doxycycline-inducible SEC14L2–enhanced green fluorescent protein (EGFP) expression plasmid (top panel). Dose-dependent effect of doxycycline (Dox) on SEC14L2–EGFP expression (middle panel) and S52/SG-neo replication (bottom panel) is shown in two single cell clones. EMCV IRES,

encephalomyocarditis internal ribosomal entry site; CFU, colony-forming units per 100,000 transfected cells. rtTA3, reverse tetracycline-controlled transactivator 3. **d**, The colonies obtained for 1 μg ml^{-1} doxycycline in **c** were pooled and passaged in G418-free medium in the presence or absence of doxycycline. After 2 weeks, SEC14L2 protein was undetectable in cells passaged without doxycycline (top panel). These cells died when subjected to G418 selection (bottom panel). Bar graphs show means ± s.d. from at least duplicate experiments.

electroporated into SEC14L2-expressing HDFR cells (SEC14L2/HDFR), varying numbers of positive cells were visible after 6 days (Extended Data Fig. 4a, b). These cells had readily detectable levels of HCV RNA that could be suppressed by treatment with daclatasvir, a specific antiviral targeting NS5A (Fig. 3a), confirming authentic HCV replication. As expected, no evidence of viral replication was seen in control HDFR cells.

For stable replication assays, we inserted a blasticidin S deaminase (BSD) cassette into full-length HCV genomes and introduced them into SEC14L2/Huh-7.5 cells by electroporation. Selection with blasticidin resulted in large numbers of colonies that were positive for viral RNA and protein (Fig. 3b, c). Inoculation of naive SEC14L2/Huh-7.5 cells with culture medium from these colonies and subsequent selection with blasticidin yielded varying numbers of colonies that were positive for viral RNA and protein, indicating production of infectious virus. Incubation of cells with anti-CD81 antibody before inoculation reduced colony formation (Extended Data Fig. 4c–e). Thus, SEC14L2 supported transient and stable replication of HCV genomes lacking adaptive mutations.

Past attempts to infect cultured cells with clinical HCV isolates have largely remained unsuccessful. Patient sera are infectious when inoculated into chimpanzees or human-liver-chimaeric mice[13,14], but do not result in productive infection in hepatoma cells. To test if SEC14L2 facilitates infection with HCV from infected sera, we inoculated SEC14L2/HDFR cells with six high-titre HCV sera of genotypes 1a, 1b and 3a. Large numbers of cells with nuclear RFP were observed for genotype 3a, followed by genotype 1b (Fig. 3d, bottom panel). All serum samples yielded detectable levels of HCV RNA that were

sensitive to treatment with daclatasvir (Fig. 3d, top panel). In contrast, control cells did not support HCV RNA replication.

A number of experiments were undertaken to examine SEC14L2-specific determinants important for its HCV replication enhancing effects. Cells expressing murine SEC14L2 were highly permissive to wild-type HCV (Extended Data Fig. 5a). Naturally occurring splice isoforms of SEC14L2, encoding shorter than the full-length proteins, did not facilitate HCV replication (Extended Data Fig. 5b, c). Two other SEC14L2-related proteins have been described in humans: SEC14L3 and SEC14L4 (ref. 11), none of which could be detected in Huh-7.5 cells (Extended Data Fig. 5d). While SEC14L3 allowed low level HCV replication, SEC14L4 had no effect (Extended Data Fig. 5d).

SEC14L2 has three domains: an amino-terminal CRAL-TRIO domain (amino acids 12–60) that binds small lipophilic molecules[15], a SEC14-like domain (amino acids 76–245) that contains a hydrophobic ligand-binding pocket, and a carboxy-terminal GOLD domain (amino acids 300–380) that is postulated to facilitate protein–protein interactions. To test the importance of these domains in facilitating HCV replication, we constructed various deletion mutants (Extended Data Fig. 5e) and SEC14L2–SEC14L4 chimaeras (Extended Data Fig. 5g). Neither the deletion mutants nor any of the chimaeric constructs conferred HCV permissiveness to Huh-7.5 cells (Extended Data Fig. 5f, h), suggesting that multiple SEC14L2-specific determinants are required for HCV replication.

In terms of mechanism, yeast two-hybrid, co-immunoprecipitation and subcellular fractionation did not reveal a physical interaction between SEC14L2 and HCV replicase components (Extended Data Fig. 6a–c). Since SEC14L2 is implicated in the regulation of the
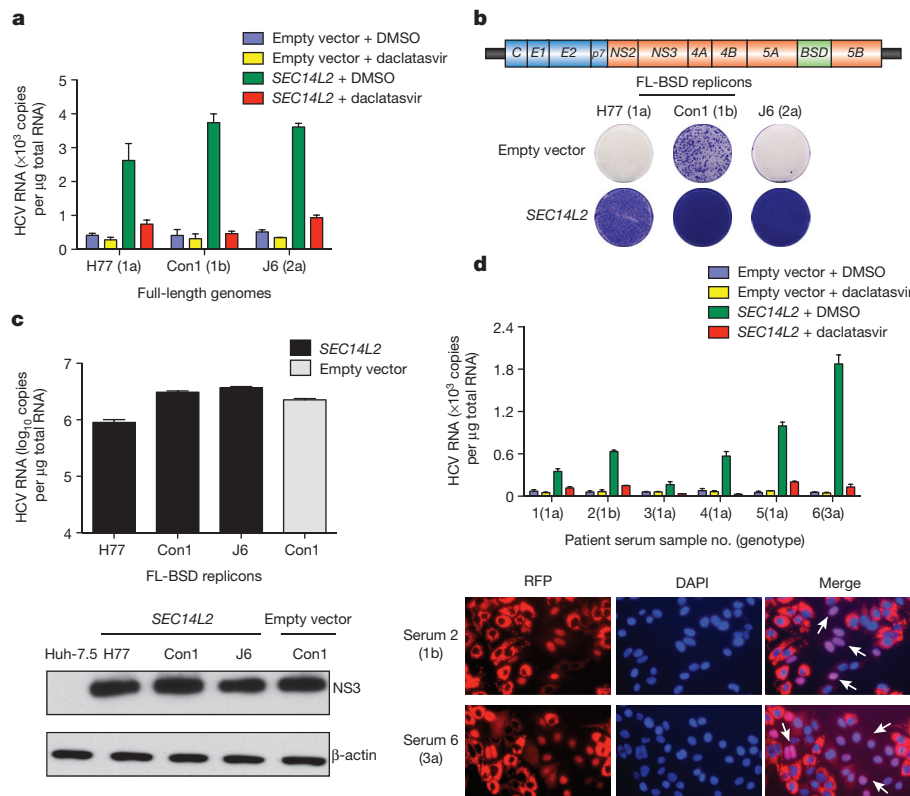

Figure 3 | SEC14L2 allows replication of full-length HCV genomes. a, HCV RNA levels 6 days after electroporation of the indicated full-length HCV genomes in HDFR reporter cells transduced with empty vector or *SEC14L2*. DMSO or daclatasvir (10 nM) was added 3 days before cell harvest. **b,** Cell colonies obtained for the indicated blasticidin-selectable, full-length (FL-BSD) HCV genomes in Huh-7.5 cells transduced or not to express SEC14L2. **c,** The cell colonies obtained in **b** were pooled and HCV RNA and protein abundance was determined by reverse transcription followed by quantitative

PCR (qRT–PCR; top panel) and immunoblotting with anti-NS3 antibody (bottom panel). **d,** Direct infection of SEC14L2-expressing HDFR reporter cells with 6 serum samples from HCV patients. The top panel shows HCV RNA levels. The bottom panel shows the nuclear localization of RFP in individual cells infected with serum 2 (genotype 1b) and serum 6 (genotype 3a). White arrows indicate cells with nuclear RFP. Bar graphs show means ± s.d. from at least duplicate experiments.

phosphoinositide 3-kinase (PI3K)/Akt signalling pathway[11,16], cholesterol synthesis[17,18], and vitamin E metabolism[16,19], we hypothesized that modulation of these pathways might be involved in facilitating HCV RNA replication. As previously reported for prostate cancer cells[16], SEC14L2 expression mildly reduced Akt phosphorylation in Huh-7.5 cells (Extended Data Fig. 7a); yet suppression of PI3K/Akt pathway by SEC14L2-independent methods failed to support HCV replication (Extended Data Fig. 7b–d).

SEC14L2 expression transiently stimulates cholesterol synthesis in rat hepatoma cells[18]. However, we did not see increased cholesterol levels in Huh-7.5 cells stably expressing SEC14L2 (data not shown). Furthermore, a SEC14L2 mutant lacking cholesterolgenic activity[18] efficiently supported HCV replication (Extended Data Fig. 7e, f). In contrast, a SEC14L2 deletion mutant capable of stimulating cholesterolgenic enzymes in rat hepatoma cells failed to confer HCV permissiveness to Huh-7.5 cells (1–383 mutant[18] in Extended Data Fig. 5e, f). Taken together, SEC14L2 phenotype cannot be attributed to stimulation of the cholesterol synthesis pathway.

We next examined the role of SEC14L2 as a vitamin E (VE; tocopherol)-binding protein[16,19]. Lipid-soluble antioxidants such as VE inhibit lipid peroxidation and enhance HCV replication[20]. As described for prostate cancer cells[16], SEC14L2 enhanced the accumulation of VE in Huh-7.5 cells (Fig. 4a). Interestingly, HCV isolates resistant to lipid peroxidation (JFH-1, Jc1, H77D, TNcc) and not enhanced by VE[20] were insensitive to SEC14L2 expression. In contrast, replication of a lipid peroxidation-sensitive clone, H77S.3 (ref. 20), was enhanced to similar levels by VE and SEC14L2 (Fig. 4b and Extended Data Fig. 8a, b). SEC14L2 and VE did not show an additive effect on H77S.3 replication (Fig. 4c, d), suggesting an overlapping mechanism. The replication enhancing effect of other lipophilic anti-oxidants was also markedly suppressed in the presence of SEC14L2 (Extended Data Fig. 8c–f).

SEC14L2 also reduced the inhibitory effect of lipophilic oxidants on HCV replication, but not that of interferon-alpha or cyclosporine A (Fig. 4e and Extended Data Fig. 8g–j), indicating selective protection against lipid peroxidation. Consistent with this hypothesis, SEC14L2 expression lowered the intracellular abundance of malondialdehyde, a secondary product of lipid peroxidation (Fig. 4f). Furthermore, SEC14L2 phenocopied the reported effects of other lipid peroxidation inhibitors such as VE and sphingosine kinase inhibitor[20] by enhancing the 50% effective concentration (EC$_{50}$) of NS3 protease (Danoprevir) and NS5B polymerase (2′CMeA) inhibitors against H77S.3, but not against lipid peroxidation-resistant Jc1 (Extended Data Fig. 8k–l).

SEC14L2 no longer supported HCV replication when cells were cultured in medium containing fetal bovine serum (FBS) that had been depleted of lipid-soluble materials including VE (delipidated FBS) (Fig. 4g, h), an effect that was reversed by VE supplementation (Fig. 4h), suggesting that the antioxidative effect of SEC14L2 is mediated, at least in part, through VE. Since delipidated FBS lacks many other lipids, we cannot entirely rule out the involvement of other potential SEC14L2 ligands in HCV replication. Recently, sphingosine kinase-2 (SPHK2) was also shown to promote lipid peroxidation and inhibit replication of non-JFH1 isolates[20]. However, neither messenger RNA nor protein levels of SPHK2 were diminished by SEC14L2 expression. In addition, inhibition of SPHK2 with sphingosine kinase inhibitor or by stable expression of dominant-negative SPHK2(G212D) (ref. 21) did not allow HCV RNA replication in Huh-7.5 cells (data not shown), suggesting that the SEC14L2 effect is not mediated by inhibiting SPHK2.

Collectively, these data imply that SEC14L2 promotes HCV replication by enhancing VE-mediated inhibition of lipid peroxidation. As expected, RNA viruses insensitive to lipid peroxidation[20] did not respond to SEC14L2 expression (Extended Data Fig. 9). Interestingly, although VE supplementation significantly enhanced SEC14L2-mediated replicon colony formation (Extended Data Fig. 10), it failed to support colony formation on its own (data not
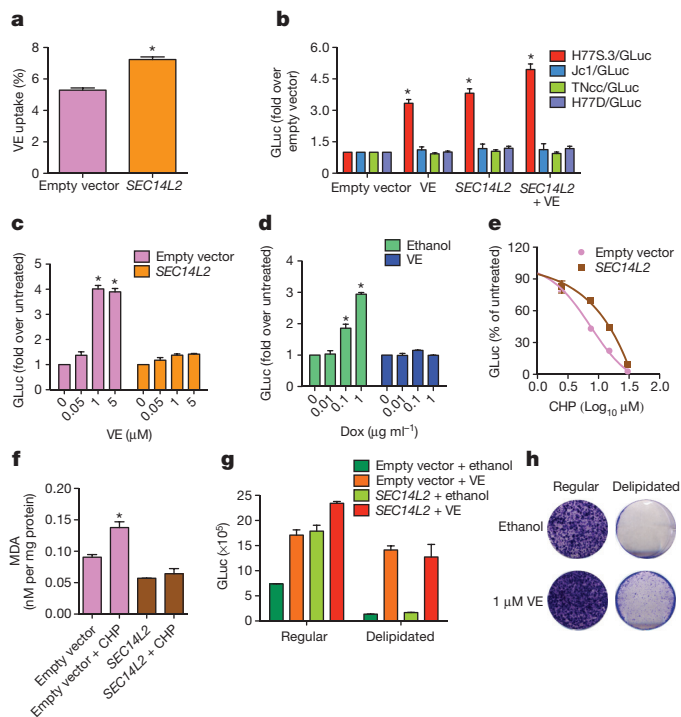

Figure 4 | SEC14L2 promotes HCV replication by enhancing the antioxidant effect of VE. a, Effect of SEC14L2 expression on VE accumulation in Huh-7.5 cells. Results represent the mean ± s.d. from 8 adjacent wells treated with radioactively labelled VE for 4 h. **b**, Effect of VE (1 μM) and SEC14L2 on replication of the indicated HCV genomes. **c**, Dose–response effects of VE on replication of H77S.3/GLuc in empty-vector- and SEC14L2-expressing Huh-7.5 cells. GLuc, *Gaussia princeps* luciferase. **d**, Dose–response effects of SEC14L2 expression on H77S.3/GLuc replication in SEC14L2–EGFP clone 1 cells (described in Fig. 2c) in presence or absence of VE. **e**, Dose–response effect of cumene hydroperoxide (CHP), a lipophilic oxidant, on replication of H77S.3/GLuc in empty-vector- and SEC14L2-expressing Huh-7.5 cells. **f**, Intracellular abundance of malondialdehyde (MDA), an end product of lipid peroxidation, in empty-vector- and SEC14L2-expressing Huh-7.5 cells incubated for 16 h with 15 μM CHP. **g**, **h**, H77S.3/GLuc replication (**g**) and S52/SG-neo colony formation (**h**) in empty-vector- and SEC14L2-expressing Huh-7.5 cells grown in regular or delipidated medium. Since delipidated serum is suboptimal for hepatoma cell proliferation, the colonies obtained were small and slow growing. Results represent means ± s.d. from two (**d**, **f**) or three (**b**, **c**, **e**, **g**) biological replicates. *P < 0.01 by one-tailed, Mann–Whitney test (**a**, **b**, **c**, **d**) or one-tailed, paired *t*-test (**f**).

shown), suggesting that SEC14L2 might have additional functions. SEC14L2-mediated enrichment of VE at specific sites may be required to support stable HCV replication, but further studies will be required to elucidate the precise mechanism by which SEC14L2 facilitates persistent replication of wild-type HCV.

This study represents an important step forward in developing a culture system where HCV can be propagated without the need for adaptive mutations. All human hepatoma cell lines that we tested became permissive to wild-type HCV genomes when engineered to express SEC14L2. Most remarkably, these cells also supported HCV replication directly from patient sera. The low levels of transient replication observed in these cells may reflect the inherently low replication potential of the wild-type HCV isolates or lack of additional host factors in Huh-7.5 cells. Moreover, although shRNA screening did not yield hits in our hands, the expression of restriction factors cannot be entirely ruled out. SEC14L2-expressing Huh-7.5 cells provide an excellent platform to test these possibilities. We anticipate that the ability of SEC14L2 to enable replication of non-adapted HCV in cell culture will open up new avenues for studying HCV biology, including drug resistant clinical variants that are emerging even in the new era of highly effective antiviral therapies[22].

1. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974 (2000).
2. Saeed, M. *et al.* Efficient replication of genotype 3a and 4a hepatitis C virus replicons in human hepatoma cells. *Antimicrob. Agents Chemother.* **56**, 5365–5373 (2012).
3. Kato, T, *et al.* Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* **125**, 1808–1817 (2003).
4. Mohd Hanafiah, K., Groeger, J., Flaxman, A. D. & Wiersma, S. T. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* **57**, 1333–1342 (2013).
5. Gottwein, J. M. *et al.* Novel infectious cDNA clones of hepatitis C virus genotype 3a (strain S52) and 4a (strain ED43): genetic analyses and *in vivo* pathogenesis studies. *J. Virol.* **84**, 5277–5293 (2010).
6. Murayama, A. *et al.* The NS3 helicase and NS5B-to-3′X regions are important for efficient hepatitis C virus strain JFH-1 replication in Huh7 cells. *J. Virol.* **81**, 8030–8040 (2007).
7. Murayama, A. *et al.* RNA polymerase activity and specific RNA structure are required for efficient HCV replication in cultured cells. *PLoS Pathog.* **6**, e1000885 (2010).
8. Bukh, J. *et al.* Mutations that permit efficient replication of hepatitis C virus RNA in Huh-7 cells prevent productive replication in chimpanzees. *Proc. Natl Acad. Sci. USA* **99**, 14416–14421 (2002).
9. Allen-Baume, V., Segui, B. & Cockcroft, S. Current thoughts on the phosphatidylinositol transfer protein family. *FEBS Lett.* **531**, 74–80 (2002).
10. Aravind, L., Neuwald, A. F. & Ponting, C. P. Sec14p-like domains in NF1 and Dbl-like proteins indicate lipid regulation of Ras and Rho signaling. *Curr. Biol.* **9**, R195–R197 (1999).
11. Kempná, P. *et al.* Cloning of novel human SEC14p-like proteins: ligand binding and functional properties. *Free Radic. Biol. Med.* **34**, 1458–1472 (2003).
12. Jones, C. T. *et al.* Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system. *Nature Biotechnol.* **28**, 167–171 (2010).
13. Farci, P. *et al.* Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc. Natl Acad. Sci. USA* **93**, 15394–15399 (1996).
14. Washburn, M. L. *et al.* A humanized mouse model to study hepatitis C virus infection, immune response, and liver disease. *Gastroenterology* **140**, 1334–1344 (2011).
15. Panagabko, C. *et al.* Ligand specificity in the CRAL-TRIO protein family. *Biochemistry* **42**, 6467–6474 (2003).
16. Ni, J. *et al.* Tocopherol-associated protein suppresses prostate cancer cell growth by inhibition of the phosphoinositide 3-kinase pathway. *Cancer Res.* **65**, 9807–9816 (2005).
17. Mokashi, V., Singh, D. K. & Porter, T. D. Supernatant protein factor stimulates HMG-CoA reductase in cell culture and *in vitro*. *Arch. Biochem. Biophys.* **433**, 474–480 (2005).
18. Mokashi, V. & Porter, T. D. Supernatant protein factor requires phosphorylation and interaction with Golgi to stimulate cholesterol synthesis in hepatoma cells. *Arch. Biochem. Biophys.* **435**, 175–181 (2005).
19. Neuzil, J., Dong, L. F., Wang, X. F. & Zingg, J. M. Tocopherol-associated protein-1 accelerates apoptosis induced by alpha-tocopheryl succinate in mesothelioma cells. *Biochem. Biophys. Res. Commun.* **343**, 1113–1117 (2006).
20. Yamane, D. *et al.* Regulation of the hepatitis C virus RNA replicase by endogenous lipid peroxidation. *Nature Med.* **20**, 927–935 (2014).
21. Yoshimoto, T. *et al.* Positive modulation of IL-12 signaling by sphingosine kinase 2 associating with the IL-12 receptor beta 1 cytoplasmic region. *J. Immunol.* **171**, 1352–1359 (2003).
22. Poveda, E. *et al.* Update on hepatitis C virus resistance to direct-acting antiviral agents. *Antiviral Res.* **108**, 181–191 (2014).

METHODS

HCV constructs. H77/SG-neo (genotype 1a), Con1/SG-neo (genotype 1b), S52/SG-neo (genotype 3a), ED43/SG-neo (genotype 4a), and SA1/SG-neo (genotype 5a) have been described[1,2,23,24]. To generate J6/SG-neo, the region of J6 full-length genome encoding the structural proteins was replaced with a cassette containing the *NPTII* gene and the EMCV IRES. Con1/SG-Feo, J6/SG-Feo, S52/SG-Feo, ED43/SG-Feo, and SA1/SG-Feo were generated by replacing the *NPTII* gene of neo replicons with a chimaeric gene encoding a fusion protein of firefly luciferase and *NPTII*, as described elsewhere[2]. A high copy number derivative of full-length H77 (pUC/H77FL-X) and its replication-defective mutant (pUC/H77FL-X pol−), referred to here as H77/AAG) have been described[25]. Molecular clones of Con1 and J6 have been previously described and shown to be infectious in chimpanzees[8,26]. We generated a high copy number derivative of Con1 consensus cDNA in pUC19 vector backbone, pUC/Con1FL-X, that has a 5′ T7 promoter and a unique XbaI run off site for production of full-length RNA transcripts. To generate blasticidin-selectable full-length replicons from these clones (H77/FL-BSD, Con1/FL-BSD, and J6/FL-BSD), we cloned a blasticidin s deaminase (*bsd*) gene flanked by the NS3/4A protease cleavage site between NS5A and NS5B. The cell culture-adapted subgenomic replicons from genotype 1a (H/SG-neo(I), H/SG-neo(D+I), H/SG-neo(L+8)), 1b (Con1/SG-neo(SI), Con1/SG-neo(PS), Con1/SG-neo(RG), Con1/SG-neo(PSRG)), 3a (S52/SG-neo(I), S52/SG-neo(VI), S52/SG-neo(AII)), and 4a (ED43/SG-neo(G), ED43/SG-neo(KG), ED43/SG-neo(YG)) have been previously described[2,27,28]. H77S.3, H77S.3/GLuc2A (referred to here as H77S.3/GLuc), H77D/GLuc, and TNcc/GLuc were a gift from S. M. Lemon[20,29]. JFH-1, J6/JFH1, and Jc1FLAG2(p7-nsGLuc2A) (referred to here as Jc1/GLuc) are previously described[30–32].

Lentiviral pseudoparticles. To generate pseudoparticles for delivery of cDNA library or individual open reading frames (ORFs), 4×10^6 Lenti-X 293T cells (Clontech) in 10-cm dishes were co-transfected with plasmids encoding (1) pLOC provirus (2) HIV-1 gag-pol/delta neo (see section Plasmids) and (3) VSV-G in a ratio of 1/0.8/0.2, respectively. To generate pseudoparticles for delivery of shRNA library or individual shRNAs, 4×10^6 Lenti-X 293T cells were co-transfected with plasmids encoding (1) pGIPZ provirus (2) pMD.G and (3) pCMVR8.91 in a ratio of 1/0.5/0.5. A total of 12 μg RNA was transfected using 24 μl jetPEI DNA transfection reagent (Polyplus Transfection, France). Transfections were carried out for 6 h, followed by a medium change to DMEM containing 10% FBS. Culture medium was collected at 48 h and 72 h, pooled, filtered through a 0.45-μm filter, and stored at −80 °C. Target cells seeded into 6-well plates at a density of 4.5×10^5 cells per well were transduced with pseudoparticles by spinoculation at 1,000*g* for 45 min at 37 °C in a medium containing 10% FBS, 20 mM HEPES, and 4 μg ml^{-1} polybrene.

shRNA/cDNA screening. For shRNA screening, Huh-7.5 cells were transduced with the pooled pGIPZ lentiviral shRNA library targeting 18,000 genes with ~60,000 shRNAs. Briefly, Huh-7.5 cells (1.8×10^7) were infected to achieve an average representation of 100 cells per shRNA. Transductions were carried out at a low transduction efficiency of 0.3 to reduce the probability of multiple integrants per cell. After 48 h, shRNA-containing cells were selected with puromycin (5 μg ml^{-1}) for 7 days. We prepared three independent cell populations transduced with the shRNA library and one cell population transduced with the control shRNA. For cDNA screening, Huh-7.5 cells were transduced with a pool of pLOC lentiviruses containing ~7,000 sequence-verified human cDNAs. Briefly, Huh-7.5 cells (2.1×10^6) were transduced to achieve a low transduction efficiency of 0.3 and an average representation of 100 cells per cDNA. After 48 h, cDNA-expressing cells were selected with blasticidin (5 μg ml^{-1}) for 10 days. We generated three independent cell populations transduced with the cDNA-containing lentiviruses and one control cell population transduced with the RFP-containing lentiviruses. The puromycin- and blasticidin-resistant cell populations were electroporated with the wild-type subgenomic replicons derived from genotype 3a (S52/SG-neo), genotype 4a (ED43/SG-neo), or as a negative control with a replication defective RNA (S52/SG-neo GNN), and selected with G418 (500 μg ml^{-1}) for 3 weeks.

Identification of integrated cDNAs. Genomic DNA was extracted from 45 replicon colonies using DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer's instructions. We subjected 20 ng DNA to PCR amplification using Expand High Fidelity PCR System (Roche) and a pair of primers complementary to the universal sequences flanking cDNAs (forward primer 5′-CTCCATAGA AGACACCGAC-3′, reverse primer 5′-GGTTCATTAGCTAGCAACCACT-3′). Touchdown PCR conditions were used to avoid non-specific bands: initial denaturation at 94 °C for 2 min, 35 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at temperature starting from 62 °C and decreasing by 0.5 °C every cycle until it reached 56 °C, and extension for 4 min at 72 °C. The resulting PCR products were resolved on a 1% agarose gel. Surprisingly, despite earlier transduction at low multiplicity, multiple bands were observed for most colonies. However, 3 colonies yielded a single PCR product encoding the same gene: *SEC14L2*.

We then designed *SEC14L2*-specific primers spanning exon–exon junctions to examine the integration of this cDNA in the remaining 42 cell colonies.

Cells and reagents. Huh-7, Huh-7.5, Hep3B/miR122[33], 293T, and BHK cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 0.1 mM nonessential amino acids (NEAA), at 37 °C in a humidified atmosphere containing 5% CO_2. All cells used in this study were tested for and found free of mycoplasma contamination. Medium containing 500 μg ml^{-1} G418 (Sigma Aldrich) was used to select and maintain cells harbouring selectable HCV subgenomic replicons. Arrayed precision lentiORFs were purchased from GE Healthcare (http://dharmacon.gelifesciences.com/mammalian-orf/precision-lentiorfs/) and pooled as previously described[34]. The GIPZ Human Whole Genome shRNA library (no. RHS6037) was also from GE Healthcare. The following primary antibodies were used in this study: SEC14L2 mouse monoclonal antibody clone 4H2 (Origene, MD), SEC14L2 rabbit polyclonal antibody (Acris Antibodies, CA), SEC14L3 rabbit polyclonal antibody (Abcam, MA), SEC14L4 rabbit polyclonal antibody (Creative BioMart, NY), HCV NS5A mouse monoclonal antibody clone 9E10[31], HCV NS3 mouse monoclonal antibody clone 2E3 (a gift from BioFront Technologies, FL), GFP rabbit polyclonal antiserum[35], pan-Akt rabbit monoclonal antibody clone C67E7 (Cell Signaling Technology), Phospho-Akt (Ser473) rabbit monoclonal antibody clone D9E (Cell Signaling Technology), PTEN mouse monoclonal antibody clone 6H2.1 (Millipore, MA), CD81 human antibody clone JS-81 (BD Biosciences, CA), calnexin rabbit polyclonal antibody (Abcam, MA), and β-actin HRP-conjugated mouse monoclonal antibody clone AC-40 (Sigma). α-tocopherol, γ-tocopherol, tocopherol succinate, docosahexaenoic acid, and 5′ azacytidine were purchased from Sigma-Aldrich. D-alpha tocopheryl quinone, and cumene hydroperoxide were from Santa Cruz Biotechnology. Linoleic acid was from Cayman Chemical. 2′CMeA and daclatasvir were from Merck Research Laboratories (PA, USA) and Selleck Chemicals (TX, USA), respectively. [^{14}C]*RRR*-α-tocopherol was a generous gift from D. Manor.

Plasmids. Human *SEC14L2* or RFP was cloned in pLOC plasmid (http://lentivirus.imb.uq.edu.au/downloads/pdfs/LOC_Technical_Manual.pdf) to obtain pSEC14L2/BlastR or pRFP/BlastR, respectively. *BlastR* gene from pSEC14L2/BlastR and pRFP/BlastR was replaced with *PuroR* to obtain the puromycin-selectable plasmids, pSEC14L2/PuroR and pRFP/PuroR, respectively. pSEC14L2–EGFP/BlastR was generated by removing the TurboGFP tag from pSEC14L2/BlastR and fusing the C-terminal end of *SEC14L2* in-frame with *EGFP*. Plasmids containing dominant negative *Akt1* (pLNCX myr HA Akt1; Addgene plasmid 9005), dominant negative p85 subunit of *PI3K* (pWZL-neo delta-p85; Addgene plasmid 10888), *SPHK2* (pDONR223-SPHK2; Addgene plasmid 23714), and constitutively active *FOXO3a* (HA-FOXO3a TM; Addgene plasmid 1788) were purchased from Addgene. Dominant-negative SPHK2(G212D) was generated by PCR-based mutagenesis to substitute aspartic acid for glycine at amino acid position 212. Expression plasmids for *SEC14L4* and mouse *Sec14l2* were from Origene. The SEC14L3 expression plasmid was obtained from CCSB-Broad lentiviral expression library (Thermo Scientific). The open reading frames (ORFs) from these plasmids were transferred to pLOC lentiviral vector (Thermo Scientific) for stable expression in Huh-7.5 cells. Akt shRNA plasmids (clones NM_005163.1628s1c1 and NM_005163.x-642s1c1) were purchased from Sigma Aldrich. The lentivirus packaging HIV-1 gag-pol plasmid was sequenced by the primer-walking method and modified by removing the NPTII gene to obtain a neomycin-sensitive derivative, HIV-1 gag-pol/delta neo.

Doxycycline-inducible SEC14L2 expression system. Tet-On pTRIPZ vector (http://dharmacon.gelifesciences.com/uploadedfiles/resources/ptripz-inducible-lentiviral-manual.pdf) was engineered to express the EGFP or SEC14L2–EGFP fusion protein under doxycycline-inducible promoter, resulting in the generation of pEGFP^{ind} and pSEC14L2–EGFP^{ind}, respectively. Lentivirus pseudoparticles were generated as described above and used to infect Huh-7.5 cells at a low multiplicity of 0.3 to reduce the probability of getting multiple integrants per cell. The untransduced cell population was eliminated by selection with puromycin (5 μg ml^{-1}) for 7 days. To obtain single cell clones, GFP-negative cells were sorted by flow cytometry at one cell per well in 96-well plates and grown to confluence. After testing multiple cell clones for absence of background GFP expression and responsiveness to doxycycline, two cell clones, SEC14L2–EGFP clone 1 and SEC14L2–EGFP clone 2 were selected for subsequent experiments.

Viruses and plaque formation assays. The following reporter viruses were used for FACS analysis: yellow fever virus (YFV)-Venus[12] (derived from YF17D-5′C25Venus2AUbi), dengue virus (DENV)-GFP[36] (generated from IC30P-A, a full-length infectious clone of strain 16681), Sindbis virus (SINV)-GFP[37] (derived from pS300/pS300-GFP), and Ross River virus (RRV)-GFP[38] (generated from the T48 strain of RRV). Plaque formation assays were performed by using the following non-reporter viruses: YFV (17D)[39] and DENV (16681)[40]. Infectious

titres of YFV were determined by plaque assay on 80–90% confluent monolayers of Huh-7.5 cells in 6-well plates. Briefly, cells were incubated with tenfold serial dilutions (prepared in DMEM containing 10% FBS) for 1 h at 37 °C, and overlaid with 1.2% Avicel (FMC BioPolymer, PA) in DMEM containing 10% FBS. After 96 h, the monolayers were fixed with 7% formaldehyde and plaques visualized by crystal violet staining. DENV titres were determined by plaque assay on 80–90% confluent monolayers of BHK cells in 12-well plates. Briefly, cells were incubated with tenfold serial dilutions (prepared in FBS-free MEM) for 2 h at 32 °C, and overlaid with 2.25% Avicel in MEM containing 5% FBS. Cells were further incubated for 5 days at 37 °C, followed by fixation with 7% formaldehyde and staining with crystal violet.

HCV RNA synthesis and transfection of cultured cells. Plasmids carrying full-length or subgenomic HCV genomes were linearized and subjected to *in vitro* transcription using the T7 RiboMAX Express Large Scale RNA Production System (Promega, WI). Complete removal of template DNA was ensured by an additional on-column treatment with RNase-free DNase and the quality of *in vitro* transcribed RNA was examined by agarose gel electrophoresis. RNA transcripts were introduced into cells either by electroporation using a BTX ElectroSquarePorator as described previously[2], or by transfection using TransIT-mRNA transfection kit (Mirus). For transfection, cells plated in 12-well plates were grown overnight to approximately 80% confluency. The culture medium was replaced with a fresh complete growth medium 30 min before transfection. The transfection complexes, prepared by adding 1 µg of HCV RNA, 2 µl of mRNA boost reagent and 2 µl of transfection reagent to 100 µl Opti-MEM in sterile tubes, were incubated for 3 min at room temperature and added drop wise to each well.

Colony formation assay. Colony formation efficiency was measured as described previously[2]. To test replicon colony formation in cells containing doxycycline-inducible SEC14L2, cells were treated with various concentrations of doxycycline for 24 h before transfection with S52/SG-neo. Forty-eight hours later, cells were subjected to G418 selection and fed with fresh G418 and doxycycline every 2 days. The resulting cell colonies were stained with crystal violet.

Luciferase assays. Cells transfected with HCV RNA encoding *Gaussia princeps* luciferase (GLuc) were exposed to compounds 6 h after transfection. The culture medium harvested between 48–72 h after adding compounds was prepared by using 1× *Renilla* luciferase assay lysis buffer (Promega) and amounts of secreted GLuc were measured with the *Renilla* luciferase assay system (Promega) and Synergy NEO HTS Multi-Mode Microplate Reader (BioTek). To test the dose–response effects of SEC14L2 expression on H77S.3/GLuc replication in presence and absence of VE, SEC14L2–EGFP clone 1 cells were treated for 24 h with ethanol (vehicle) or 1 µM VE in presence of the indicated concentrations of doxycycline. Cells were then transfected with H77S.3/GLuc RNA for 6 h and incubated with VE and doxycycline for another 72 h before measuring the amounts of secreted GLuc.

HCV replication in delipidated medium. Huh-7.5 cells transduced with empty vector or *SEC14L2* were passaged 3 times in the culture medium supplemented with 10% of regular or delipidated FBS. These cells were then transfected with H77S.3/GLuc or S52/SG-neo for 6 h, followed by medium change to ethanol- or VE-supplemented medium (1 µM). H77S.3/GLuc replication was examined by measuring the secreted GLuc activity at 72 h. For S52/SG-neo colony formation, the transfected cells were subjected to G418 selection at 48 h after transfection and fed with fresh G418 every 2 days. After 3 weeks of selection, the resulting cell colonies were stained with crystal violet.

Inoculation of cells with HCV-infected sera. Six high-titre serum samples (5.98×10^6 to 3.2×10^7 IU ml^{-1}), four of genotype 1a and one of each of genotypes 1b and 3a were obtained from HCV patients at the Mount Sinai Hospital, as approved by the Institutional Review Board of the Mount Sinai Hospital. Informed consent was obtained from all patients. For infection, 25 µl of patient serum diluted in 500 µl of DMEM containing 10% FBS, 1× NEAA, and 50 ng ml^{-1} EGF was added to cells grown overnight in 24-well plates. After overnight incubation with serum, cells were rinsed twice with 1× PBS and fed with 500 µl fresh medium containing 10% FBS, 1× NEAA, and 50 ng ml^{-1} EGF. Four days later, the cells were exposed to DMSO or daclatasvir (10 nM) for 72 h, after which they were harvested for qPCR or fixed with 4% PFA/PBS to test the nuclear localization of RFP.

Construction of SEC14L2 isoforms 2 and 3. SEC14L2 has three splice isoforms. Isoform 1, identified in the screen, is 403 amino acids long, comprised of 12 exons. Isoform 2 is 392 amino acids that share the N-terminal 360 amino acids with isoform 1 but alternative splicing at exon 10 generates a unique C terminus. Isoform 3 is generated by skipping of exons 4 and 5, resulting in a 320 amino acid protein. We used isoform 1 as a PCR template to construct the other two isoforms. Isoform 2 was generated by multiple rounds of PCR with the following primers: 1st round (forward primer 5′-GCTCGGATCCACTAGTCCAGTG-3′, reverse primer 5′-AGGCAGAGATACTTACAGATGCCAGGATCAC-3′), 2nd round (forward primer 5′-GCTCGGATCCACTAGTCCAGTG-3′, reverse primer 5′-

TTCAAGGCATTGCCAAGGCAGAGATACTTAC-3′), 3rd round (forward primer 5′-GCTCGGATCCACTAGTCCAGTG-3′, reverse primer 5′-GAAAGCTGGACATGGGGCTTCAAGGCATTGC-3′), 4th round (forward primer 5′-GCTCGGATCCACTAGTCCAGTG-3′, reverse primer 5′-GAGGAACCTCACAGGCAGAAAGCTGGACATG-3′), 5th round (forward primer 5′-GCTCGGATCCACTAGTCCAGTG-3′, reverse primer 5′-AAAAATCCATGGAGAAGAGGAACCTCACAG-3′), and 6th round (forward primer 5′-GCTCGGATCCACTAGTCCAGTG-3′, reverse primer 5′-TAGCTAGCTCATTAACACTCAGAGCCAAAAATCCATGGAG-3′). The resulting PCR product was digested with BamHI and NheI and replaced with the corresponding fragment of pLOC-SEC14L2 isoform 1. To generate isoform 3, the coding region for amino acids 55–140 of isoform 1 was deleted by a two-step PCR. In the first step, PCR was carried out with primers 5′-GCTCGGATCCACTAGTCCAGTG-3′/5′-CTTCCTCCCCAACTTCCGGAGCATGGCCT-3′ and 5′-AGGCCATGCTCCGGAAGTTGGGGAGGAAG-3′/5′-TAGCTAGCTCATTATTTCGGGGTGCCTGC-3′. Since the 5′ end of the internal reverse primer was complementary to the 5′ end of the internal forward primer, the two PCR products obtained in the first step were pieced together by a second round of PCR with primers 5′-GCTCGGATCCACTAGTCCAGTG-3′ and 5′-TAGCTAGCTCATTATTTCGGGGTGCCTGC-3′. The resulting PCR product was digested with BamHI and NheI and used to replace the corresponding fragment of pLOC-SEC14L2 isoform 1.

Yeast two-hybrid assays. Yeast two-hybrid binding assays were performed as described previously[41], using the Matchmaker GAL4 Yeast Two Hybrid 3 system (Clontech, Mountain View, CA). *SEC14L2* was cloned into pGADT7 (activation domain) vector (Clontech) using 5′ EcoRI and 3′ BamHI cloning sites. HCV nonstructural protein encoding genes were cloned into pGBKT7 (DNA binding domain) vector (Clontech) as described previously[41]. *Saccharomyces cerevisiae* AH-109 was co-transformed with *SEC14L2* and individual viral gene-containing plasmids and selected by culturing on Sabouraud dextrose broth plates in the absence of Leu and Trp for 3 days at 30 °C. For testing protein–protein interactions, ~100 colonies were pooled, resuspended and selected on Sabouraud dextrose broth plates (minus Leu, minus Trp, minus Ade, minus His) for 3 days at 30 °C.

Co-immunoprecipitation assay. SEC14L2–EGFP/Huh-7.5 cells (1×10^7) harbouring wild-type Con1/SG-neo were harvested, washed 3 times with 1 ml of ice-cold phosphate-buffered saline (PBS), and suspended in 0.5 ml of lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$ and 1% Triton X-100) supplemented with EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland), DNase I (50 U ml^{-1}) and RNase A (10 µg ml^{-1}). Lysates were clarified by centrifugation (16,000g, 10 min, 4 °C) and incubated with the mouse monoclonal anti-NS5A antibody (9E10)[31] or the rabbit polyclonal anti-GFP antibody[42] captured on Dynabeads Protein A (Invitrogen, Carlsbad, CA) for 2 h at 4 °C. The beads were then washed 3 times in lysis buffer and the bound proteins were released by boiling in SDS–PAGE loading buffer and analysed by immunoblotting.

Subcellular fractionation. SEC14L2/Huh-7.5 cells (2×10^7) harbouring Con1/SG-neo were washed twice with PBS, scraped into PBS, and pelleted by centrifugation at 1,000g for 5 min at 4 °C. Cells were resuspended in 150 mM sucrose containing 10 mM HEPES and protease inhibitor cocktail without EDTA (Roche Diagnostic, Manheim, Germany) and lysed by cell disruption with a Parr Bomb at 900 psi. Nuclei and unbroken cells were removed by centrifugation at 1,000g for 10 min at 4 °C. The supernatant was carefully removed and overlaid onto a discontinuous sucrose gradient (460 µl of 56% sucrose, 920 µl of 50% sucrose, 1.38 ml of 45% sucrose, 2.3 ml of 40% sucrose, 2.3 ml of 35% sucrose, 1.38 ml of 30% sucrose, 460 µl of 20% sucrose, and 1 ml of 8% sucrose containing 10 mM HEPES, protease inhibitor and trypsin inhibitor) and centrifuged at 200,000g for 18 h at 4 °C in a SW41Ti rotor. 12 fractions were collected from the top and SEC14L2, calnexin, and HCV NS5A were detected by immunoblotting.

Vitamin E uptake assay. Stable Huh-7.5 cell populations expressing SEC14L2 or RFP were seeded into 24-well plates at a density of 5×10^4 cells per well and grown overnight. [^{14}C]*RRR*-α-tocopherol pre-mixed with FBS was added to the culture medium. After 4 h of incubation at 37 °C, the media was removed, the cells washed with DMEM and then lysed in 20 mM HEPES (pH 7.4), 1 mM EDTA, 150 mM NaCl, 1% NP40, 20 mM sodium fluoride, 20 mM β-glycerophosphate, 1 mM sodium vanadate, and 200 µM PMSF. The radioactivity in cell lysate and culture medium was measured in a scintillation counter, and vitamin E (VE) accumulation was calculated as follows:

Tocopherol accumulation (%) = $CPM_{cells}/(CPM_{cells}+CPM_{medium}) \times 100$

The values obtained were then normalized to the total protein and presented as VE uptake percentage.

Measurement of lipid peroxidation. 1.5×10^6 cells seeded into 100-mm dishes were treated overnight with cumene hydroperoxide (15 µM), and malondialdehyde (MDA) levels were measured by a previously described classical method[43]. Briefly, cell monolayers were washed with 1× PBS, scraped into 1× PBS and

pelleted by centrifugation at 1,000g for 5 min at 4 °C. The cell pellet was resuspended in 110 μl PBS and incubated on ice for 30 min. 10 μl of cell suspension was taken for protein measurement and the rest was used for MDA quantification. Briefly, 100 μl of cell suspension was mixed with 100 μl of 10% SDS, followed by addition of freshly made 100 μl of 0.75% thiobarbituric acid, 30% trichloroacetic acid, and 0.5 N HCl and incubated at 95 °C for 15 min. The reaction was stopped by incubating the mixture on ice for 15 min, followed by centrifugation for 5 min at 13,500g. 100 μl of each sample was transferred to a clear bottom black 96-well microplate. The fluorescence at 550 nm excitation/530 nm was measured by reference to a standard curve prepared with 0 to 10 μM MDA.

Statistical analysis. No statistical methods were used to predetermine sample size. Statistical significance was tested using non-parametric, one-tailed Mann–Whitney test (for two-way comparisons) and one-tailed paired t-test (for multiple comparisons) in Prism v5.0 (GraphPad). A value of $P < 0.05$ was considered significant. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

23. Lohmann, V. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285,** 110–113 (1999).
24. Wose Kinge, C. N. *et al.* Hepatitis C virus genotype 5a subgenomic replicons for evaluation of direct-acting antiviral agents. *Antimicrob. Agents Chemother.* **58,** 5386–5394 (2014).
25. Liehl, P. *et al.* Host-cell sensors for *Plasmodium* activate innate immunity against liver-stage infection. *Nature Med.* **20,** 47–53 (2014).
26. Yanagi, M., Purcell, R. H., Emerson, S. U. & Bukh, J. Hepatitis C virus: an infectious molecular clone of a second major genotype (2a) and lack of viability of intertypic 1a and 2a chimeras. *Virology* **262,** 250–263 (1999).
27. Lohmann, V., Korner, F., Dobierzewska, A. & Bartenschlager, R. Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *J. Virol.* **75,** 1437–1449 (2001).
28. Blight, K. J., McKeating, J. A., Marcotrigiano, J. & Rice, C. M. Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *J. Virol.* **77,** 3181–3190 (2003).
29. Shimakami, T. *et al.* Protease inhibitor-resistant hepatitis C virus mutants with reduced fitness from impaired production of infectious virus. *Gastroenterology* **140,** 667–675 (2011).
30. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Med.* **11,** 791–796 (2005).
31. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309,** 623–626 (2005).
32. Marukian, S. *et al.* Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* **48,** 1843–1850 (2008).
33. Fukuhara, T. *et al.* Expression of microRNA miR-122 facilitates an efficient replication in nonhepatic cells upon infection with hepatitis C virus. *J. Virol.* **86,** 7918–7933 (2012).
34. Rodriguez-Barrueco, R., Marshall, N. & Silva, J. M. Pooled shRNA screenings: experimental approach. *Methods Mol. Biol.* **980,** 353–370 (2013).
35. Yi, Z. *et al.* Identification and characterization of the host protein DNAJC14 as a broadly active flavivirus replication modulator. *PLoS Pathog.* **7,** e1001255 (2011).
36. Schoggins, J. W. *et al.* Dengue reporter viruses reveal viral dynamics in interferon receptor-deficient mice and sensitivity to interferon effectors *in vitro. Proc. Natl Acad. Sci. USA* **109,** 14610–14615 (2012).
37. Suthar, M. S., Shabman, R., Madric, K., Lambeth, C. & Heise, M. T. Identification of adult mouse neurovirulence determinants of the Sindbis virus strain AR86. *J. Virol.* **79,** 4219–4228 (2005).
38. Shabman, R. S. *et al.* Differential induction of type I interferon responses in myeloid dendritic cells by mosquito and mammalian-cell-derived alphaviruses. *J. Virol.* **81,** 237–247 (2007).
39. Bredenbeek, P. J. *et al.* A stable full-length yellow fever virus cDNA clone and the role of conserved RNA elements in flavivirus replication. *J. Gen. Virol.* **84,** 1261–1268 (2003).
40. Kinney, R. M. *et al.* Construction of infectious cDNA clones for dengue 2 virus: strain 16681 and its attenuated vaccine derivative, strain PDK-53. *Virology* **230,** 300–308 (1997).
41. Chung, H. Y., Gu, M., Buehler, E., MacDonald, M. R. & Rice, C. M. Seed sequence-matched controls reveal limitations of small interfering RNA knockdown in functional and structural studies of hepatitis C virus NS5A-MOBKL1B interaction. *J. Virol.* **88,** 11022–11033 (2014).
42. Cristea, I. M., Williams, R., Chait, B. T. & Rout, M. P. Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* **4,** 1933–1941 (2005).
43. Pan, M. *et al.* Lipid peroxidation and oxidant stress regulate hepatic apolipoprotein B degradation and VLDL production. *J. Clin. Invest.* **113,** 1277–1287 (2004) .
44. Scheel, T. K. *et al.* Analysis of functional differences between hepatitis C virus NS5A of genotypes 1-7 in infectious cell culture systems. *PLoS Pathog.* **8,** e1002696 (2012).
45. Maqbool, M. A. *et al.* Regulation of hepatitis C virus replication by nuclear translocation of nonstructural 5A protein and transcriptional activation of host genes. *J. Virol.* **87,** 5523–5539 (2013).
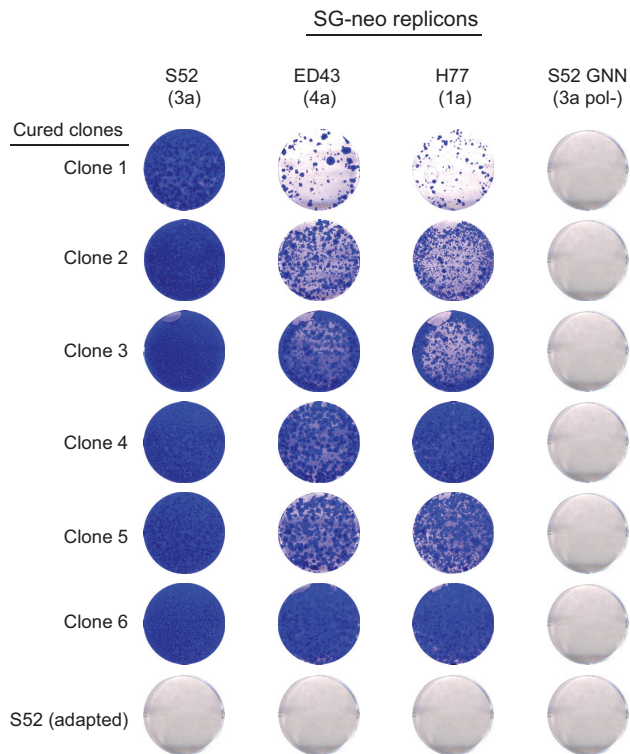
Extended Data Figure 1 | Cured replicon cell clones supported replication of wild-type HCV. Six of the 45 cell colonies obtained from cDNA screen and one control cell clone, S52 (adapted) that carried a cell culture-adapted HCV subgenomic replicon S52/SG-neo(AII), were treated for 15 days with a combination of an HCV NS5A inhibitor (daclatasvir, 1 nM) and an NS5B polymerase inhibitor (2′CMeA, 1 μM) to eliminate the replicating viral genomes. These cells were then transfected with the indicated wild-type HCV replicons and selected with G418 for 3 weeks. The resulting cell colonies were stained with crystal violet. Replication defective replicon, S52 GNN, with a catalytically inactive mutation in the NS5B polymerase served as a negative control.
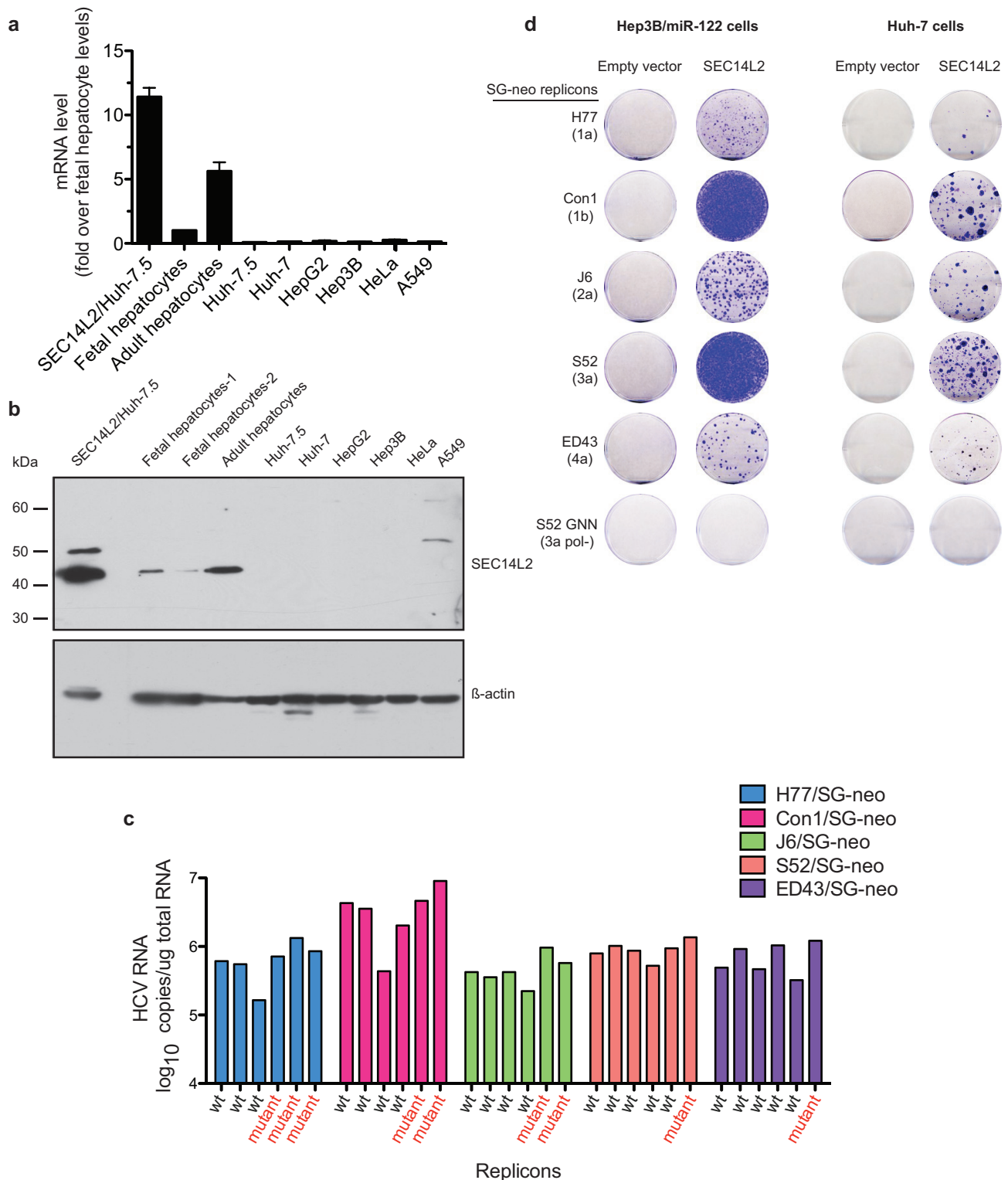
Extended Data Figure 2 | The ectopic expression of SEC14L2 confers HCV permissiveness to human hepatoma cells. a, b, *SEC14L2* is not expressed in most cell lines. **a,** *SEC14L2* mRNA levels were measured in the indicated cells by qPCR, and the values were normalized to those of the housekeeping gene, *RPS11*. Shown is the fold difference from fetal hepatocytes. The results are plotted as mean ± s.e.m. of two different cell stocks. **b,** SEC14L2 protein expression in the indicated cells was determined by immunoblotting with SEC14L2 rabbit polyclonal antibody. β-actin is included as a loading control. **c,** SEC14L2-expressing Huh-7.5 cells were transfected with the indicated wild-type subgenomic replicons. After 4 weeks of selection, HCV RNA was sequenced in six individual colonies from each of the replicons. HCV RNA levels in each of these colonies were determined by qRT–PCR. **d,** Hep3B/miR-122 and Huh-7 cells transduced with *SEC14L2* or empty vector were transfected with *in vitro*-transcribed RNA from the indicated wild-type replicons and selected for 3 weeks with G418 (500 μg ml^{-1}). The resulting cell colonies were stained with crystal violet. S52 GNN with a catalytically inactive mutation in NS5B polymerase was included as a negative control.
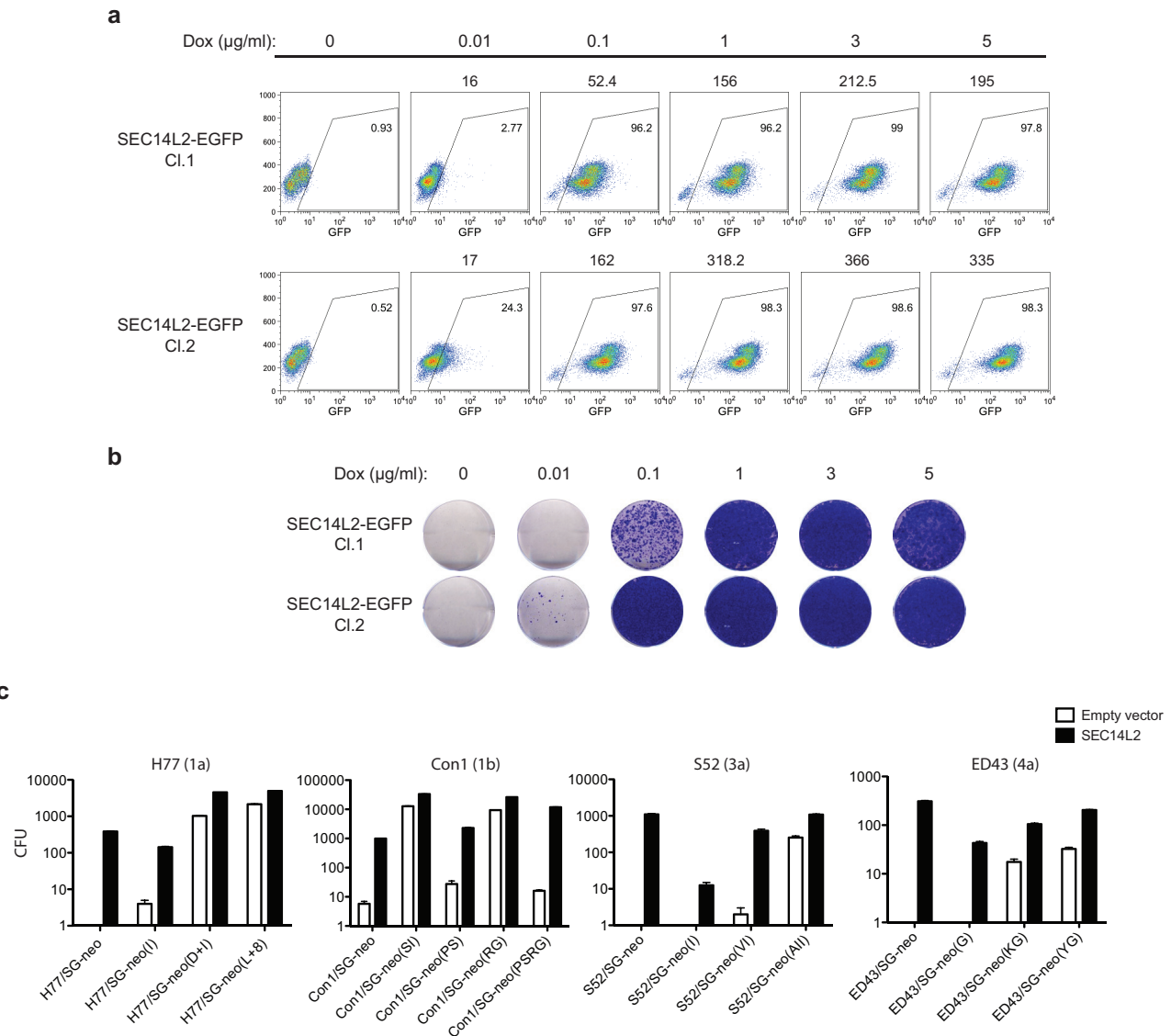
Extended Data Figure 3 | Effect of SEC14L2 expression on the replication of wild-type and cell culture-adapted HCV subgenomic replicons.
a, b, SEC14L2 expression enhances replication of wild-type HCV RNA in a dose-dependent manner. **a,** Huh-7.5 cells were transduced with lentiviruses encoding SEC14L2–EGFP fusion protein under a doxycycline-inducible promoter and flow cytometry was performed to obtain single-cell clones. Two cell clones selected for downstream analysis were treated with the indicated concentrations of doxycycline for 24 h, followed by flow cytometry to determine the number of EGFP-positive cells. Mean fluorescence intensities of EGFP are shown at the top of each box. **b,** S52/SG-neo colony formation in cells described in **a**. The results were confirmed by two independent transfections. **c,** SEC14L2 expression enhances replication of cell culture-adapted HCV replicons to varying extents. Colony formation efficiency of the indicated subgenomic replicons in empty-vector- and SEC14L2-expressing Huh-7.5 cells is plotted as CFU per 100,000 transfected cells. Results represent mean ± s.e.m. from two independent transfections.
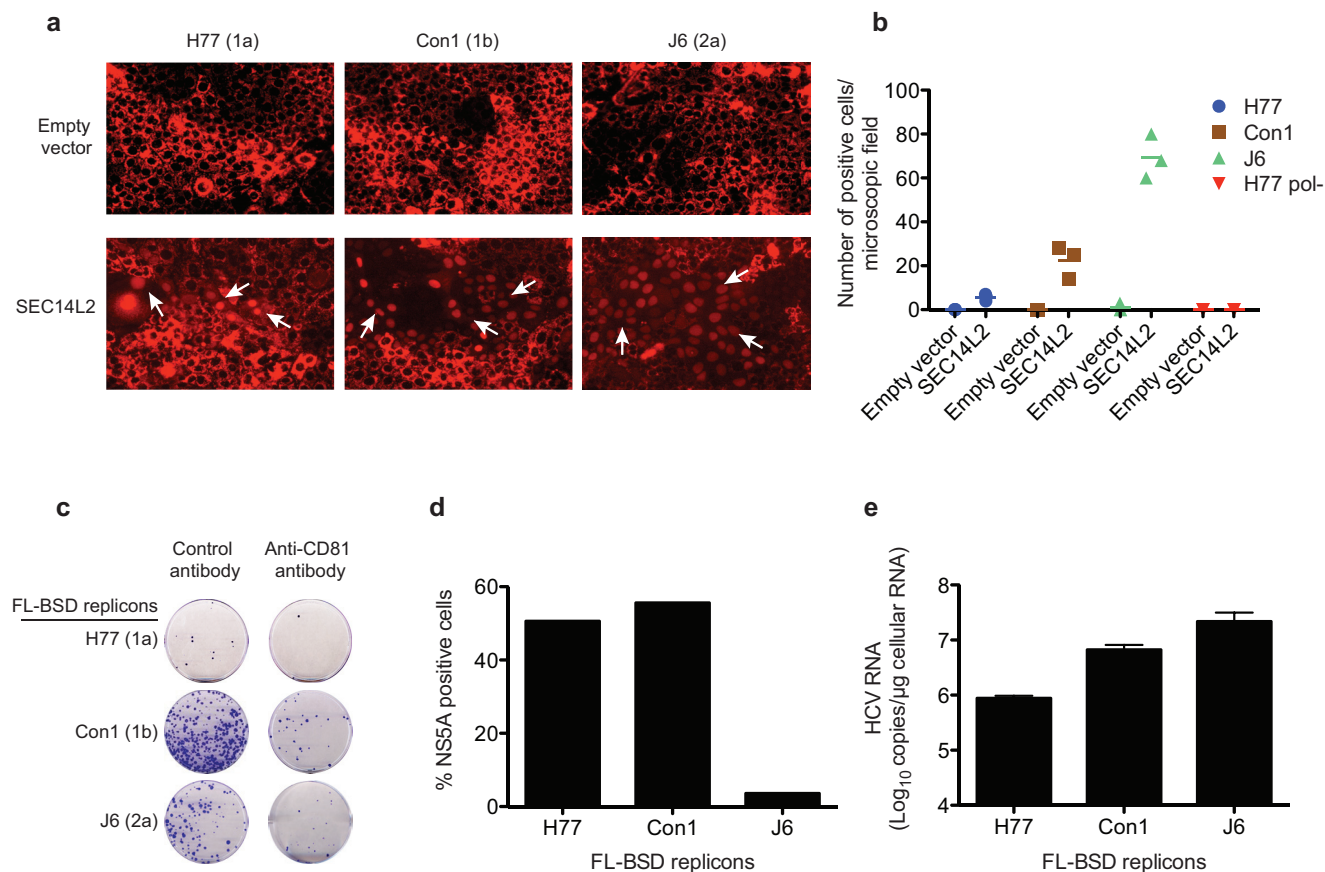
Extended Data Figure 4 | SEC14L2 expression enables replication of wild-type full-length HCV genomes and production of low levels of infectious virus particles. a, HDFR reporter cells transduced with *SEC14L2* or empty vector were transfected with *in vitro* transcribed RNA from H77, Con1, and J6 full-length genomes. Live cell images were captured 6 days after transfection. White arrows indicate the cells with nuclear RFP. **b**, The numbers of cells exhibiting nuclear RFP were counted in 3 random microscopic fields at day 6 post-transfection. H77 pol−, with a catalytically inactive mutation in NS5B polymerase, was used as a negative control. **c–e**, Infectious virus particles are produced from SEC14L2/Huh-7.5 cells harbouring selectable full-length HCV genomes. **c**, Huh-7.5 cells stably expressing SEC14L2 were incubated at 37 °C for 30 min with anti-CD81 antibody or control IgG, and inoculated with the culture medium from SEC14L2/Huh-7.5 cells harbouring blasticidin-selectable, full-length (FL-BSD) H77, Con1, and J6 genomes, described in Fig. 3b. After 72 h, selection with blasticidin (2.5 µg ml^{-1}) was imposed and the colonies obtained after 3 weeks were stained with crystal violet. **d**, The cell colonies obtained in **c** were pooled and stained with anti-NS5A antibody (9E10 clone). The percentages of positive cells from two independent experiments are plotted. As previously described[44], 9E10 antibody did not detect NS5A from the J6 isolate. **e**, The cell colonies obtained in **c** were pooled and HCV RNA levels were measured by qPCR. The values are plotted as mean ± s.e.m. of two independent experiments.
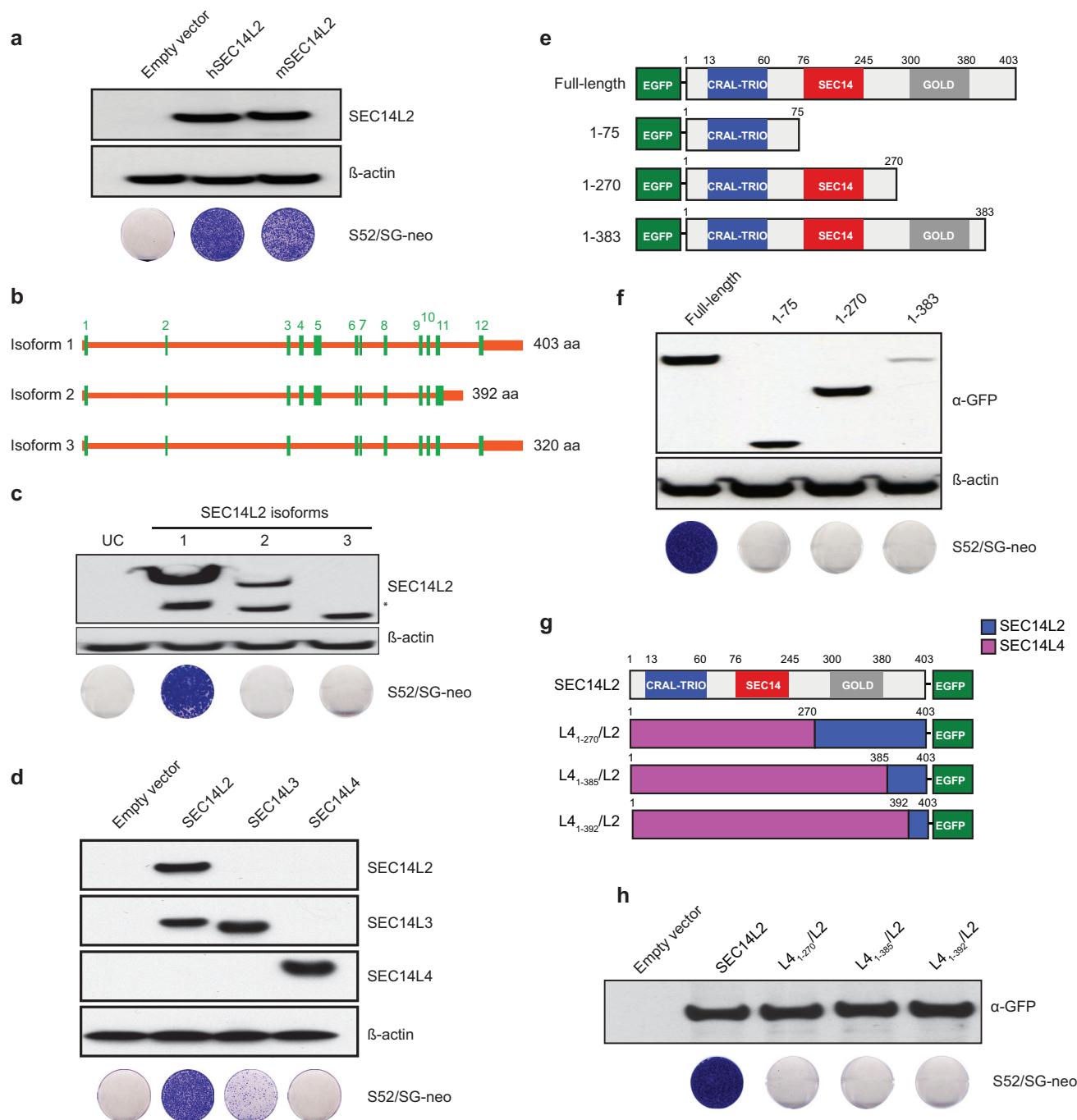
Extended Data Figure 5 | Only full-length SEC14L2 supports replication of wild-type HCV. a, Huh-7.5 cells stably expressing human or murine *SEC14L2* (93% amino acid identity) were lysed and protein expression was confirmed by immunoblotting. These cells were then transfected with S52/SG-neo and selected with G418. The resulting cell colonies were stained with crystal violet. **b, c**, Only isoform 1 of SEC14L2 supports HCV RNA replication. The *SEC14L2* gene is comprised of 12 exons and results in 3 alternatively spliced transcript variants encoding 3 different protein isoforms. Isoform 1 was identified in cDNA screening. **b**, Schematic representation of SEC14L2 isoforms. Coding exons are shown as green blocks and the amino acid length of each protein is shown on right. **c**, Cell lysates from Huh-7.5 cells stably expressing 3 SEC14L2 isoforms were analysed by 4–12% SDS–PAGE and immunoblotting was performed with SEC14L2 mouse monoclonal antibody. The bands highlighted with asterisks might be a cleavage product of SEC14L2. UC, untransduced cells. These cells were transfected with S52/SG-neo and selected with G418 for 3 weeks. The resulting cell colonies were stained with crystal violet. **d**, SEC14L3 and SEC14L4 (86% and 80% amino acid similarity to SEC14L2, respectively)[11] are not expressed in Huh-7.5 cells. Cell lysates from Huh-7.5 cells stably expressing SEC14L2, SEC14L3, or SEC14L4 were analysed by 4–12% SDS–PAGE followed by immunoblotting with the indicated antibodies

(the signal generated by SEC14L3 antibody in SEC14L2-expressing cells most likely reflects the cross-reactivity of SEC14L3 antibody). These cells were then tested for their ability to support replication of S52/SG-neo. **e, f**, Deletion mutants of SEC14L2 do not support HCV RNA replication. **e**, Schematic representation of N-terminal EGFP-tagged SEC14L2 deletion mutants. **f**, Cell lysates from Huh-7.5 cells stably expressing the full-length SEC14L2 and the deletion mutants were analysed by 4–12% SDS–PAGE followed by immunoblotting with the indicated antibodies. These cells were then transfected with S52/SG-neo and selected with G418 for 3 weeks. The resulting cell colonies were stained with crystal violet. **g, h**, As N-terminal deletion mutants of SEC14L2 were unstable in Huh-7.5 cells (they formed protein aggregates), we generated chimaeric constructs by fusing C-terminal ends of SEC14L2 with the corresponding N-terminal sequences from SEC14L4 and tested their ability to support HCV replication. **g**, Schematic representation of C-terminal EGFP-tagged chimaeric constructs. **h**, Huh-7.5 cells stably expressing the indicated chimaeric constructs were lysed and protein expression was confirmed by immunoblotting with anti-GFP antibody. The cells were then transfected with S52/SG-neo and selected with G418 for 3 weeks.
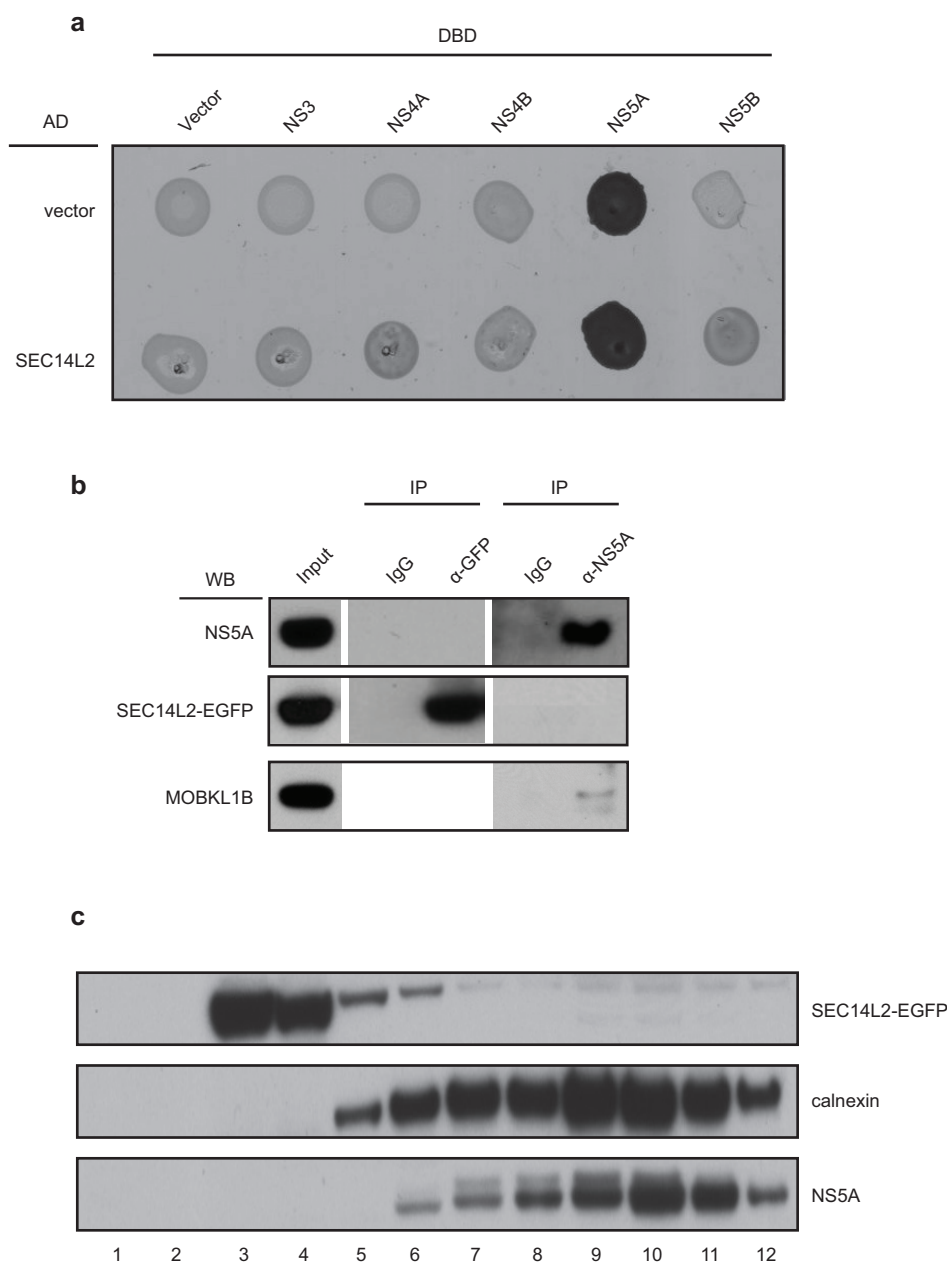
Extended Data Figure 6 | SEC14L2 does not interact with HCV non-structural proteins under the tested conditions. a, Yeast two-hybrid assay was performed to examine direct interaction between SEC14L2 and HCV non-structural proteins. SEC14L2-AD (GAL4 activation domain) fusion or control AD vector was co-expressed with DBD (DNA binding domain) fusion of the individual HCV non-structural proteins or control DBD vector and tested for positive yeast two-hybrid interactions under selective nutritional conditions (lacking leucine, tryptophan, histidine, and adenosine). The strong signal obtained for NS5A most likely reflects the intrinsic trans-activating activity of NS5A[45]. **b**, Co-immunoprecipitation did not reveal binding of SEC14L2 with HCV NS5A protein. As two-hybrid assay (**a**) did not yield unambiguous results on interaction between SEC14L2 and NS5A, we employed co-immunoprecipitation assay to probe binding between these proteins. Huh-7.5 cells stably expressing SEC14L2–EGFP and harbouring wild-type Con1/SG-neo replicon were lysed and subjected to immunoprecipitation with anti-NS5A antibody, anti-EGFP antibody or control immunoglobulins (IgG). The bound proteins were analysed by immunoblotting. Endogenous MOBKL1B, a previously described binding partner of the NS5A protein[41], served as a positive control. **c**, Subcellular fractionation of SEC14L2–EGFP/Huh-7.5 cells harbouring wild-type Con1/SG-neo replicon did not reveal obvious co-fractionation of SEC14L2 and HCV NS5A protein.
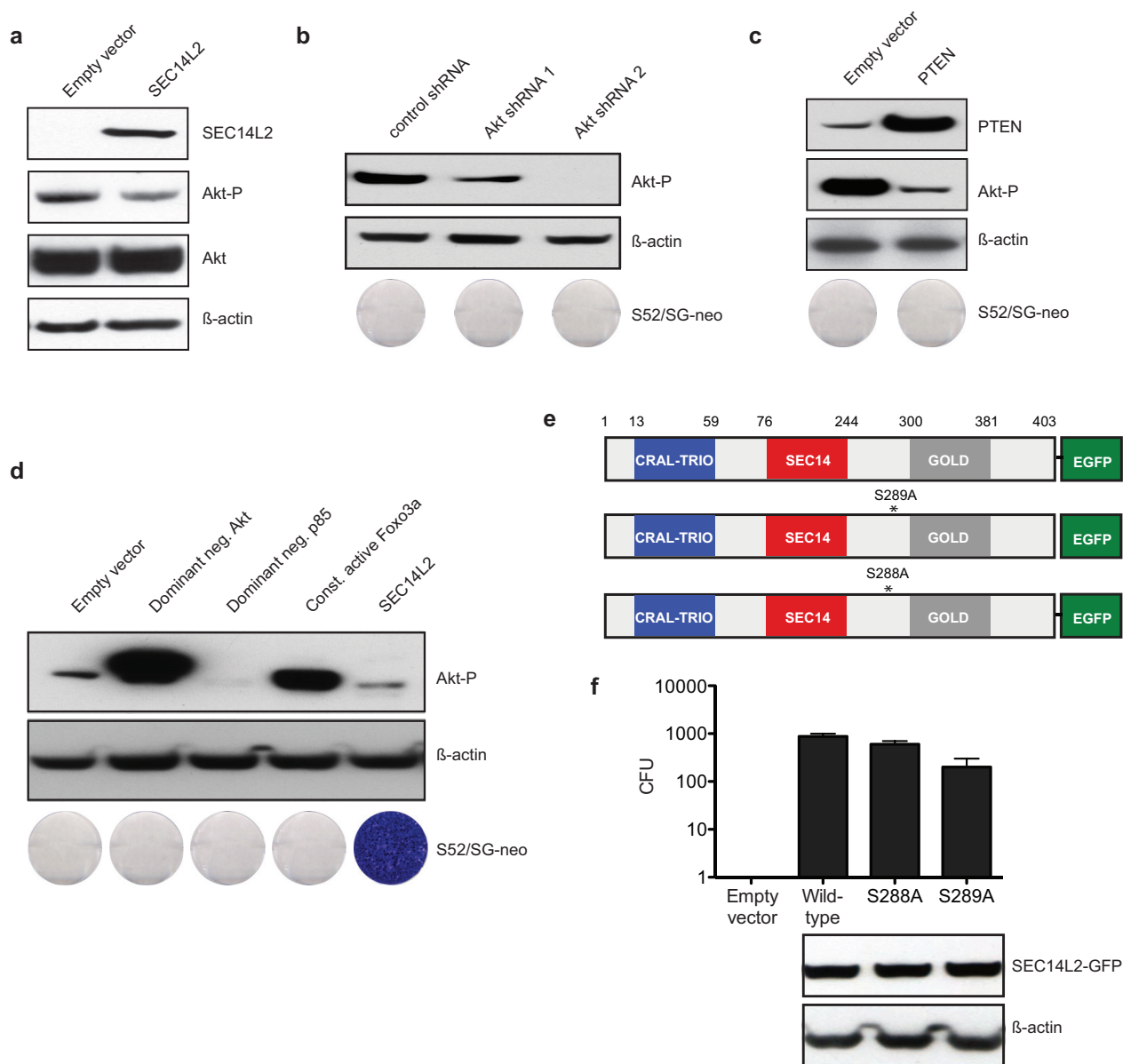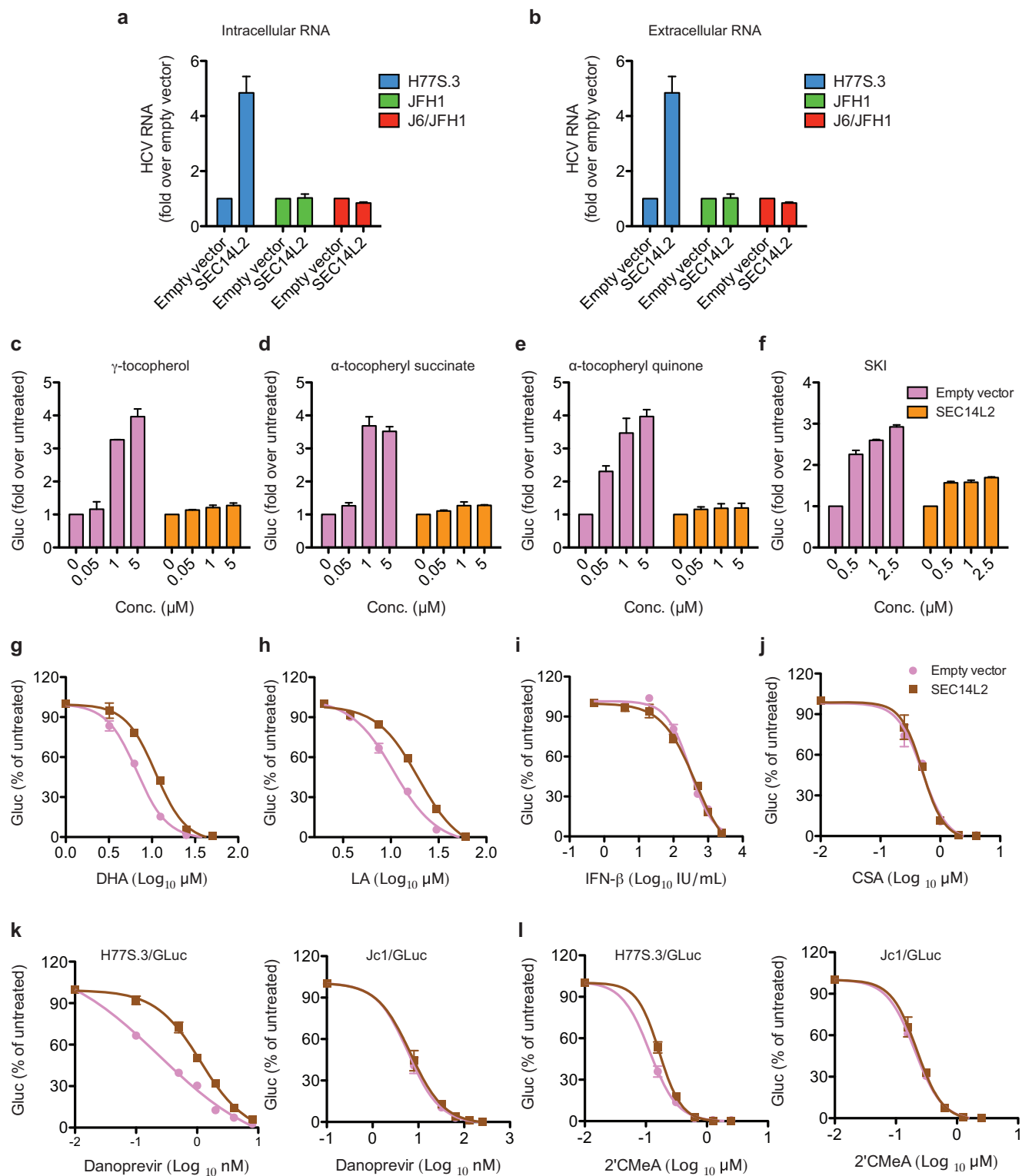
Extended Data Figure 7 | SEC14L2 does not facilitate HCV RNA replication by modulating PI3K/Akt or cholesterol synthesis pathways. a–d, Downregulation of PI3K/Akt pathway does not support HCV replication. **a,** Akt phosphorylation in control and SEC14L2-expressing Huh-7.5 cells was analysed by immunoblotting; β-actin is included as a loading control. **b,** Stable knockdown of Akt in Huh-7.5 cells with two different shRNAs did not facilitate G418 resistant colony formation by S52/SG-neo. **c,** Huh-7.5 cells were transduced to stably express PTEN, a negative regulator of PI3K pathway. Despite decreased Akt phosphorylation, these cells did not support colony formation by S52/SG-neo. **d,** Suppression of PI3K pathway in Huh-7.5 cells by stable expression of a dominant negative Akt, a dominant negative p85

subunit of PI3K, or a constitutively active FOXO3a did not render them permissive to HCV replication. Dominant negative Akt gets phosphorylated, but as it is kinase-dead, it cannot initiate the downstream signalling. Interestingly, increased Akt phosphorylation was seen in Huh-7.5 cells expressing constitutively active FOXO3a, suggesting a potential feedback mechanism. **e, f,** A SEC14L2 mutant (S289A) lacking the cholesterolgenic activity supported HCV replication. **e,** Schematic representation of carboxy-terminal EGFP-tagged SEC14L2 point mutants. S288A was generated as a negative control. **f,** The colony formation efficiency of S52/SG-neo (plotted as CFU per 100,000 transfected cells) and the expression levels of SEC14L2 mutants in Huh-7.5 are shown.

a Intracellular RNA

b Extracellular RNA

c γ-tocopherol

d α-tocopheryl succinate

e α-tocopheryl quinone

f SKI

g DHA (Log_{10} µM)

h LA (Log_{10} µM)

i IFN-β (Log_{10} IU/mL)

j CSA (Log_{10} µM)

k H77S.3/GLuc — Danoprevir (Log_{10} nM); Jc1/GLuc — Danoprevir (Log_{10} nM)

l H77S.3/GLuc — 2'CMeA (Log_{10} µM); Jc1/GLuc — 2'CMeA (Log_{10} µM)

Extended Data Figure 8 | SEC14L2 expression masks the effects of lipophilic oxidants and anti-oxidants on H77S.3/GLuc replication.
**a**, **b**, SEC14L2 enhances transient replication of H77S.3, but not that of JFH-1 or J6/JFH1. Empty-vector- and SEC14L2-expressing Huh-7.5 cells were electroporated with the indicated viral RNAs lacking GLuc insertions. **a**, Intracellular and **b**, extracellular RNA levels were measured 6 days after electroporation. Results are plotted as fold change from empty vector control. **c–f**, The pro-viral effects of various lipophilic antioxidants on H77S.3/GLuc replication were suppressed in the presence of SEC14L2. **c**, γ-tocopherol, **d**, α-tocopheryl succinate, **e**, α-tocopheryl quinone, and **f**, sphingosine kinase inhibitor (SKI) were added to Huh-7.5 cells 20 h before transfection with H77S.3/GLuc. Transfections were carried out in the fresh medium lacking these compounds. 6 h after transfection, cells were again fed with each compound and the GLuc expression was measured at 72 h post-transfection. The results are presented as fold change from untreated cells. **g–l**, SEC14L2 expression suppressed the inhibitory effect of lipophilic oxidants and direct-acting antivirals, but not that of the HCV host factor inhibitors, on H77S.3/GLuc replication. **g**, Docosahexaenoic acid (oxidant); **h**, linoleic acid (oxidant); **i**, interferon-β (IFN-β); **j**, CSA (cyclosporine A); **k**, danoprevir (NS3 protease inhibitor) and **l**, 2′CMeA (NS5B polymerase inhibitor) were added to Huh-7.5 cells 6 h after transfection with H77S.3/GLuc (and Jc1/GLuc in the case of Danoprevir and 2′CMeA) and the secreted GLuc activity at 72 h post-transfection was measured. The results are plotted as percentage of inhibition relative to the untreated cells. All results in this figure represent mean ± s.d. of two replicate experiments.
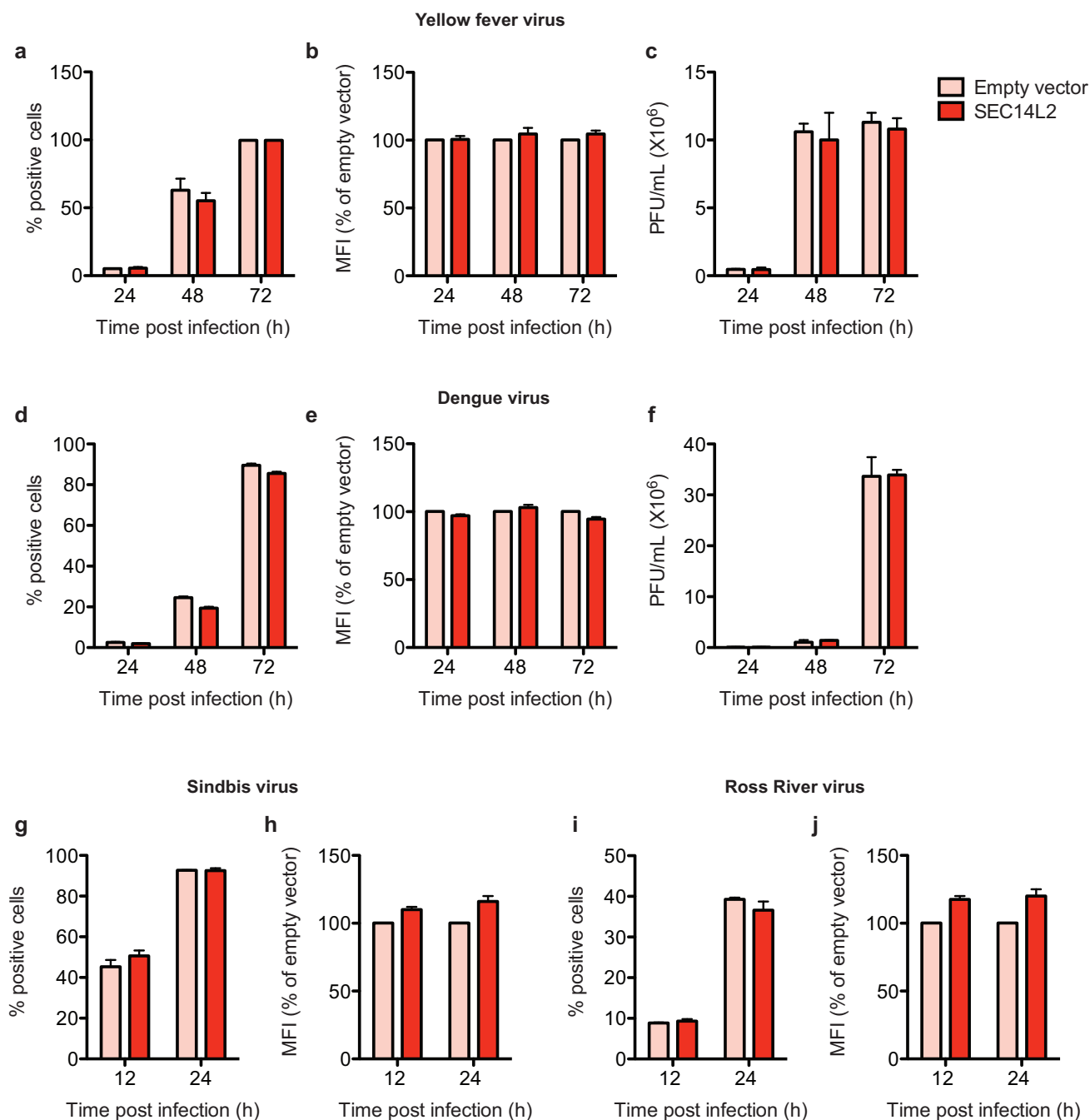
Extended Data Figure 9 | SEC14L2 expression does not enhance replication of lipid peroxidation-resistant RNA viruses. a–f, Flaviviruses, such as yellow fever virus (YFV) and dengue virus (DENV) do not respond to SEC14L2 expression. **a–e,** Empty-vector- and SEC14L2-expressing Huh-7.5 cells were infected with YFV-venus (**a, b**) or DENV-GFP (**d, e**) at a multiplicity of infection (MOI) of 0.01 and 0.1, respectively. The cells were harvested at the indicated times post-infection, followed by FACS analysis. **a, d,** The number of yellow (YFV-venus) and green (DENV-GFP) cells are plotted as percentage of positive cells and **b, e,** the mean fluorescence intensities (MFI) are presented as percentage relative to empty vector. **c, f,** Empty-vector- and SEC14L2-expressing Huh-7.5 cells were inoculated with **c,** YFV (17D) or **f,** DENV

(serotype 2 strain 16681) at an MOI of 0.5 and 0.1, respectively. The culture medium was collected at the indicated times post infection and the infectious virus titres were determined by a plaque formation assay on Huh-7.5 cells (YFV-venus) or BHK cells (DENV). The results are presented as plaque-forming units (PFU) per ml. **g–j,** Alphaviruses, such as Sindbis virus (SINV) and Ross River virus (RRV) are insensitive to SEC14L2 expression. Empty-vector- and SEC14L2-expressing Huh-7.5 cells were infected with SINV-GFP (**g, h**) or RRV-GFP (**i, j**) at an MOI of 0.05. The cells were harvested at the indicated times post-infection and FACS analysis was carried out to determine the number of green cells and the mean fluorescence intensities. All results in this figure represent mean ± s.e.m. of two replicate experiments.

a

b
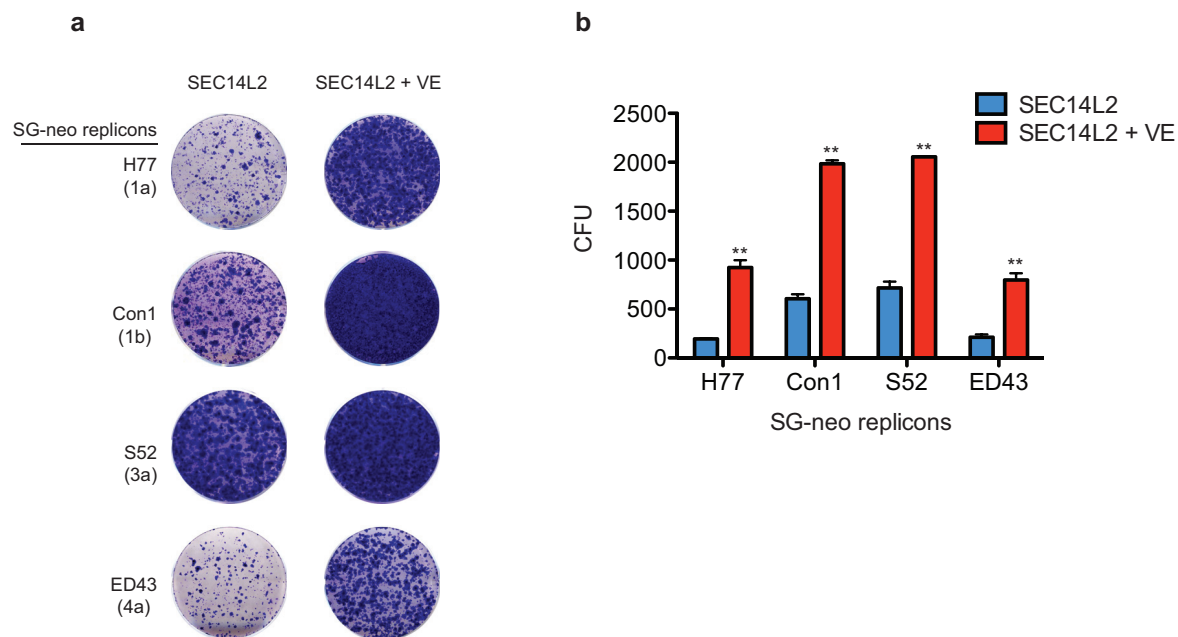
Extended Data Figure 10 | VE increased SEC14L2-mediated replicon colony formation. a, VE (α-tocopherol, 1 μM) was added to Huh-7.5 cells 20 h before transfection with the indicated wild-type HCV subgenomic replicons. Transfections were carried out for 6 h in the fresh medium lacking VE, followed by medium change to VE-containing medium. After 48 h, cells were subjected to G418 selection and fed with fresh G418 and VE every 2 days. The resulting cell colonies were stained with crystal violet. Shown are the results of one of three independent experiments. **b**, Colony formation efficiency of the indicated subgenomic replicons was measured in SEC14L2-expressing cells in the absence or presence of 1 μM VE. The results are plotted as mean ± s.e.m. of CFU per 100,000 transfected cells from three independent transfections. **$P < 0.005$ by two-tailed, paired t-test.

LETTER

Cell mixing induced by *myc* is required for competitive tissue invasion and destruction

Romain Levayer[1], Barbara Hauert[1] & Eduardo Moreno[1]

Cell–cell intercalation is used in several developmental processes to shape the normal body plan[1]. There is no clear evidence that intercalation is involved in pathologies. Here we use the proto-oncogene *myc* to study a process analogous to early phase of tumour expansion: *myc*-induced cell competition[2–7]. Cell competition is a conserved mechanism[5,6,8,9] driving the elimination of slow-proliferating cells (so-called 'losers') by faster-proliferating neighbours (so-called 'winners') through apoptosis[10] and is important in preventing developmental malformations and maintain tissue fitness[11]. Here we show, using long-term live imaging of *myc*-driven competition in the *Drosophila* pupal notum and in the wing imaginal disc, that the probability of elimination of loser cells correlates with the surface of contact shared with winners. As such, modifying loser–winner interface morphology can modulate the strength of competition. We further show that elimination of loser clones requires winner–loser cell mixing through cell–cell intercalation. Cell mixing is driven by differential growth and the high tension at winner–winner interfaces relative to winner–loser and loser–loser interfaces, which leads to a preferential stabilization of winner–loser contacts and reduction of clone compactness over time. Differences in tension are generated by a relative difference in F-actin levels between loser and winner junctions, induced by differential levels of the membrane lipid phosphatidylinositol (3,4,5)-trisphosphate. Our results establish the first link between cell–cell intercalation induced by a proto-oncogene and how it promotes invasiveness and destruction of healthy tissues.

To analyse quantitatively loser cell elimination, we performed long-term live imaging of clones showing a relative decrease of the proto-oncogene *myc* in the *Drosophila* pupal notum (Fig. 1a, b and Supplementary Video 1), a condition known to induce cell competition[3,4] in the wing disc. Every loser cell delamination was counted over 10 h and we calculated the probability of cell elimination for a given surface of contact shared with winner cells (Fig. 1c, d and Methods). We observed a significant increase of the proportion of delamination with winner–loser shared contact, whereas this proportion remained constant for control clones (Supplementary Video 2 and Fig. 1d). The same correlation was observed in *ex vivo* culture of larval wing disc (Extended Data Fig. 1 and Supplementary Video 3). Cell delamination in the notum was apoptosis dependent (Supplementary Video 4; *UAS-diap1*, 1 out of 922 cells delaminated, 4 nota) and expression of *flower*^*lose* (*fwe*^*lose*), a competition-specific marker for loser fate[12], was necessary and sufficient to drive contact-dependent delamination (Fig. 1d and Supplementary Videos 5 and 6). Moreover we confirmed[11,12] that contact-dependent death is based on the computation of relative differences of *fwe*^*lose* between loser cells and their neighbours (Extended Data Fig. 2a–e). Thus, cell delamination in the notum recapitulates features of cell competition[4,12].

This suggested that winner–loser interface morphology could modulate the probability of eliminating loser clones. Using the wing imaginal disc, we reduced winner–loser contact by inducing adhesion- or tension-dependent cell sorting[13] (Fig. 2d) and observed a significant reduction of loser clone elimination (Extended Data Figs 2f, g

and 3a–c). This rescue was not driven by a cell-autonomous effect of E-cadherin (E-cad) or active myosin II regulatory light chain (MRLC) on growth, death or cell fitness (Extended Data Fig. 3d–f) but rather by a general diminution of winner–loser contact. Competition is ineffective across the antero-posterior compartment boundary[14], a frontier that prevents cell mixing through high line tension[15]. Accordingly, there was no increase in death at the antero-posterior boundary in wing discs overexpressing *fwe*^*loseA* in the anterior compartment (Fig. 1g, h). However, reducing tension by reducing levels of myosin II heavy chains was sufficient to increase the shared surface of contact between cells of the anterior and posterior compartments (Fig. 1e, f, $P < 10^{-5}$), and induced *fwe*^*lose* death at the boundary (Fig. 1g, h). Altogether, we concluded that the reduction in surface contact between winners and losers is sufficient to block competition, which explains how compartment boundaries prevent competition.

Loser clones have been reported to fragment more often than controls[14,16], whereas winner clones show convoluted morphology[14,17], suggesting that winner–loser mixing is increased during competition. This could affect the outcome of cell competition by increasing the surface shared between losers and winners. We used clone splitting as a readout for loser–winner mixing. Two non-exclusive mechanisms can drive clone splitting: cell death followed by junction rearrangement (Fig. 2a), or junction remodelling and cell–cell intercalation independent of death (Fig. 2b). To assess the contribution of each phenomenon, we systematically counted the proportion of clones fragmented 48 h after clone induction (ACI) (Fig. 2c; see also Methods and Extended Data Fig. 4a, b). We observed a twofold increase in the frequency of split clones in losers (wild type (WT) in *tub-dmyc*) versus WT in WT controls (Fig. 2d and Extended Data Fig. 4c). Overexpressing E-cad or active myosin II was sufficient to prevent loser clone splitting, whereas blocking apoptosis or blocking loser fate by silencing *fwe*^*lose* did not reduce splitting (Fig. 2d and Extended Data Fig. 4c). Finally, the proportion of split clones was also increased for winner clones either during *myc*-driven competition (*UAS-myc, UAS-p35*; Fig. 2d and Extended Data Fig. 4c) or during Minute-dependent competition[2] (WT clones in $M^{-/+}$ background; Fig. 2d and Extended Data Fig. 4c). Altogether, this suggested that winner–loser mixing is increased independently of loser cell death or clone size (Extended Data Fig. 4d) by a factor upstream of *fwe*, and could be driven by cell–cell intercalation. Accordingly, junction remodelling events leading to disappearance of a loser–loser junction were three times more frequent at loser clone boundaries than control clone boundaries in the pupal notum (Fig. 2e, f and Supplementary Video 7, $P < 10^{-4}$). The rate of junction remodelling was higher in loser–loser junctions and in winner–winner junctions than in winner–loser junctions (Extended Data Fig. 5a, b). The preferential stabilization of winner–loser interfaces should increase the surface of contact between winner and loser cells over time. Accordingly, loser clone compactness in the notum decreased over time whereas it remains constant on average for WT clones in WT background (Fig. 2g, h; $P < 10^{-4}$). Similarly, the compactness of clones in the notum also decreased over time for conditions showing high frequency of clone splitting in the wing disc, whereas clone

[1]Institute for Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland.
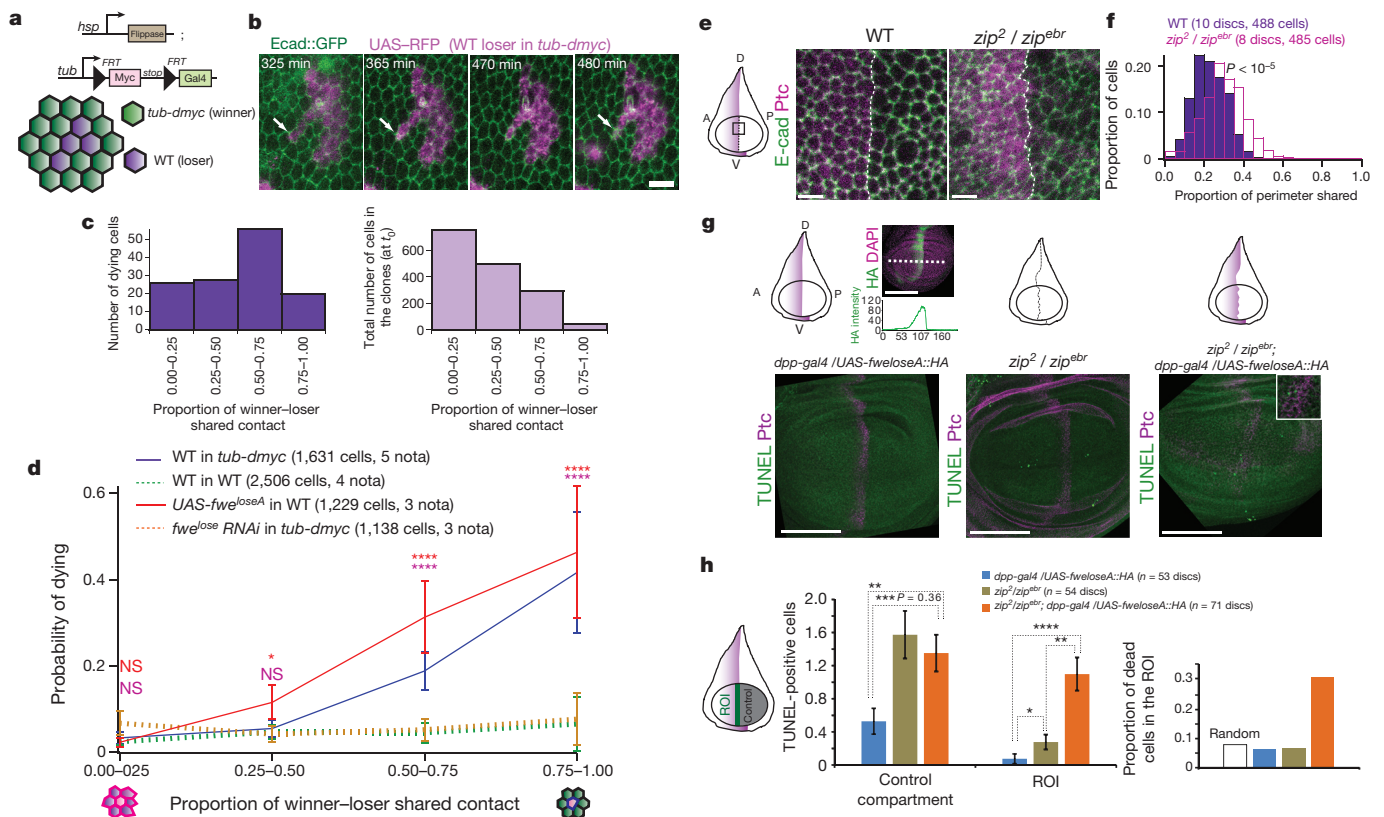
Figure 1 | Loser elimination correlates with the surface shared with winners.
a, Description of the supercompetition assay[4]. **b**, Snapshots of Supplementary Video 1 showing loser cell delamination (white arrowheads) in the pupal notum (a single layer epithelium). Scale bar, 10 μm. **c**, Distribution of the proportion of junctional perimeter of loser cells shared with winner cells 1 h before delamination (left) and in all the loser cells at t_0 (right). 0 = cell in the centre of the clone, 1 = isolated cell surrounded by winners. **d**, Probability of loser cell elimination for a given surface of contact shared with winners. Statistical tests are Fisher's exact tests performed with WT in WT. Error bars, 95% confidence intervals. NS (not significant), $P > 0.05$; $*P < 0.05$; $**P < 10^{-2}$; $***P < 10^{-3}$; $****P < 10^{-4}$. **e**, Adherens junctions (E-cad) of cells at the wing disc antero-posterior compartment boundary (white dotted line, localized with Patched (Ptc)) in WT and upon downregulation of myosin II heavy chain (zip^2/zip^{ebr}). Pictures are representative of ten and eight wing discs, respectively. A, anterior; P, posterior; D, dorsal; V, ventral. Scale bars, 5 μm. **f**, Distribution of the proportion of perimeter shared with cells across the

antero-posterior (AP) compartment boundary (P values, Mann–Whitney test). **g**, The z-projection of a wing disc overexpressing $fwe^{loseA}::HA$ in the anterior compartment (left, representative of 53 discs), a zip^2/zip^{ebr} disc (middle, representative of 54 discs) and a disc combining both (right, representative of 71 discs). Inset at right shows TdT-mediated dUTP nick end labelling (TUNEL)-positive cells at the antero-posterior boundary. Top schematic: purple, $fwe^{loseA}::HA$; black line, antero-posterior compartment boundary. Note that $fwe^{loseA}::HA$ is expressed in a graded way in the anterior compartment (top left). Scale bars, 100 μm. **h**, Average number of TUNEL-positive cells in the posterior compartment (control) and in the vicinity of the antero-posterior boundary (region of interest (ROI), left scheme). Right: ratio of average number of dead cells in the ROI over the total number of dead cells in the wing pouch. The white box is the expected ratio for a random distribution (ROI surface/total wing pouch surface). Error bars, s.e.m. P values, Mann–Whitney test. $*P < 0.05$; $**P < 10^{-2}$; $***P < 10^{-3}$; $****P < 10^{-4}$.

compactness remained constant for conditions rescuing clone splitting (compare Fig. 2d with Extended Data Fig. 5d, e). Altogether, we concluded that both Minute- and *myc*-dependent competition increase loser–winner mixing through cell–cell intercalation.

We then asked what could modulate the rate of junction remodelling during competition. The rate of junction remodelling can be cell-autonomously increased by *myc* (Extended Data Fig. 5c and Supplementary Video 8). Interestingly, downregulation of the tumour suppressor PTEN is also sufficient to increase the rate of junction remodelling[18] through the upregulation of phosphatidylinositol (3,4,5)-trisphosphate (PIP3). We reasoned that differences in PIP3 levels could also modulate junction remodelling during competition. Using a live reporter of PIP3 that could detect modulations of PIP3 in the notum (Extended Data Fig. 6a, b), we observed a significant increase of PIP3 in the apico-lateral membrane of *tub-dmyc–tub-dmyc* interfaces compared with WT–WT and WT–*tub-dmyc* interfaces (Fig. 3a, b). Moreover, increasing/reducing Myc levels in a full compartment of the wing disc was sufficient to increase/decrease the levels of phospho-Akt (a downstream target of PIP3 (ref. 19), Fig. 3c), whereas *fwe^{loseA}* overexpression had no effect (Extended Data Fig. 6c). Similarly, levels of phospho-Akt were relatively higher in

WT clones than in the surrounding $M^{-/+}$ cells (Extended Data Fig. 6d). Thus differences in PIP3 levels might be responsible for winner–loser mixing. Accordingly, reducing PIP3 levels by overexpressing a PI3 kinase dominant negative (PI3K-DN) or increasing PIP3 levels by knocking down PTEN (*UAS-pten RNAi*) were both sufficient to induce a high proportion of fragmented clones (Fig. 3d, e) and to reduce clone compactness over time in the notum (Extended Data Fig. 5d, e), whereas increasing PIP3 in loser clones was sufficient to prevent cell mixing (Fig. 3d, e). Moreover, abolishing winner–loser PIP3 differences through larval starvation[20] (Extended Data Fig. 6e–g) prevented loser clone fragmentation (Extended Data Fig. 6h, i), the reduction of clone compactness over time in the notum (Extended Data Fig. 5d, e) and could rescue WT clone elimination in *tub-dmyc* background (Extended Data Fig. 6j, k). We therefore concluded that differences in PIP3 levels are necessary and sufficient for loser–winner mixing and required for loser cell elimination.

We then asked which downstream effectors of PIP3 could affect junction stability. A relative growth decrease can generate mechanical stress[21,22] that can be released by cell–cell intercalation. Accordingly, growth reduction through Akt downregulation is sufficient to increase clone splitting (Extended Data Fig. 7a, b) and could contribute to loser
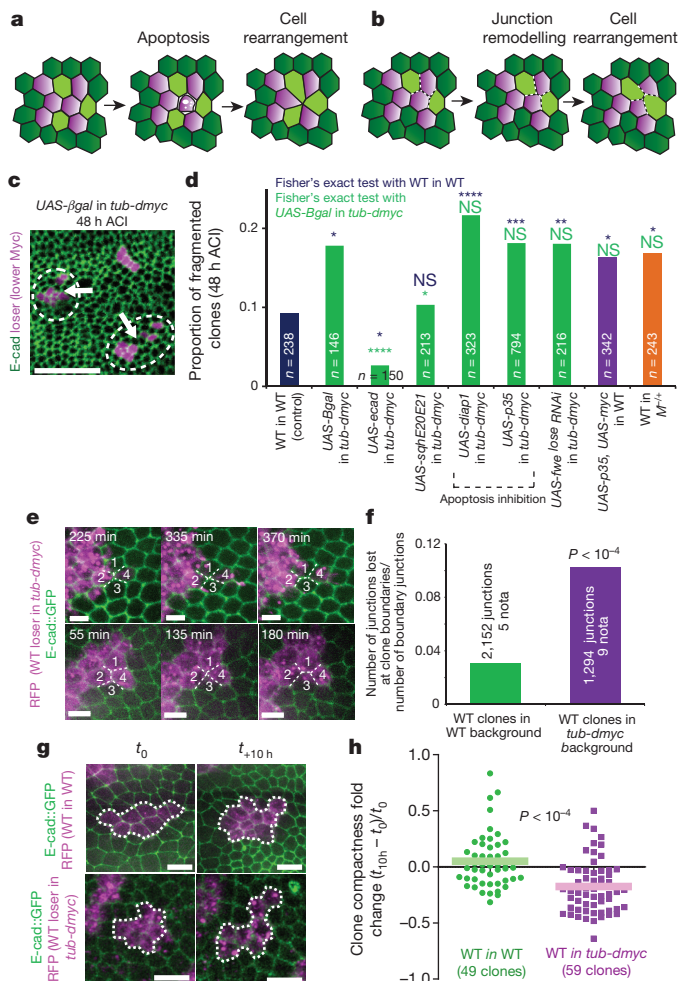
Figure 2 | Winner–loser mixing is induced by junction remodelling and cell–cell intercalation. **a**, Clone fragmentation induced by cell death and the subsequent cell rearrangement **b**, Clone fragmentation induced by cell–cell intercalation. **c**, Top left, supercompetition assay in the wing disc 48 h ACI (loser cells, purple). White arrowheads, fragmented clones (representative of 146 clones analysed). Scale bar, 25 μm. **d**, Proportion of fragmented clones (images in Extended Data Fig. 4c); n = number of clones (test: Fisher's exact tests with WT in WT in blue, and UAS-βgal in tub-$dmyc$ in green; NS, $P > 0.05$; $*P < 0.05$; $**P < 10^{-2}$; $***P < 10^{-3}$; $****P < 10^{-4}$). **e**, Junction remodelling events in loser clones in the pupal notum 48 h ACI leading to clone splitting (top) or persistent disappearance of a loser–loser junction (bottom). Scale bars, 5 μm. **f**, Proportion of red fluorescent protein (RFP)–RFP junctions at clone boundaries disappearing through remodelling after 12 h in the pupal notum (P values, Fisher's exact test). **g**, WT in WT clone and WT in tub-$dmyc$ clone in the notum at time 0 and 10 h later (48 h ACI, 20 h after pupae formation, representative of 49 and 59 clones, respectively). Scale bars, 5 μm. **h**, Evolution of single clone compactness over time in the notum (($t_{10\,h} - t_0)/t_0$, see Methods). One dot or one square represents one clone; bars are averages. P values, Mann–Whitney test.

clone splitting. However, Akt is not sufficient to explain winner–loser mixing because, unlike PIP3, increasing Akt had no effect on clone splitting (Extended Data Fig. 7a, b). PIP3 could also modulate junction remodelling through its effect on cytoskeleton[23] and the modulation of intercellular adhesion or tension[1]. We could not detect obvious modifications of E-cad, MRLC or Dachs (another regulator of tension[24]) in loser cells (Extended Data Fig. 7c–f). However, we observed a significant reduction of F-actin levels and a reduction of actin turnover/polymerization rate in loser–loser and loser–winner junctions in the notum (Fig. 4a, b, $P < 10^{-5}$; see also Extended Data Fig. 8 and Supplementary Video 9). Similarly, modifying Myc levels in a full wing
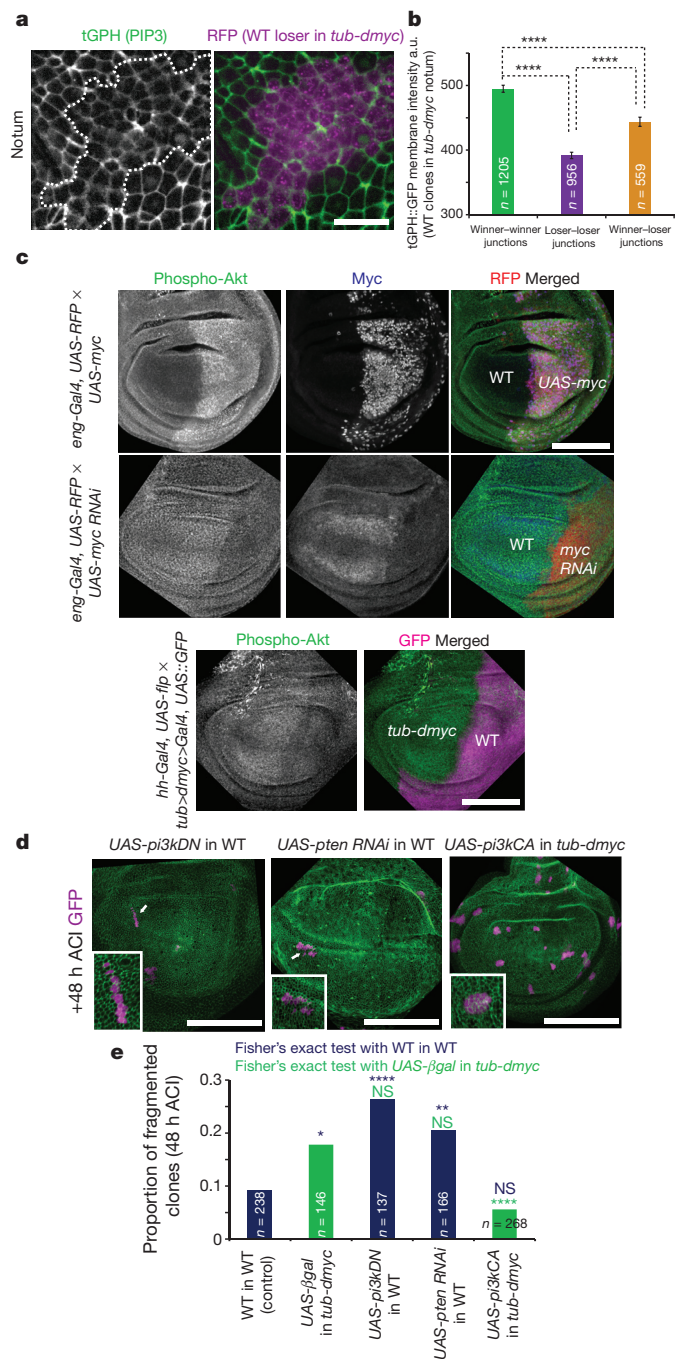

Figure 3 | Differences in PIP3 induce loser–winner mixing. **a**, The z-projection of tubulin–green fluorescent protein (GFP)–pleckstrin homology (tGPH) (PIP3) around the apico-lateral domain of loser clones (purple) in the pupal notum 48 h ACI (white dotted lines are clone boundaries, representative of 44 clones). Scale bar, 10 μm. **b**, Mean membrane intensity of tGPH; n, number of junctions. $****P < 10^{-4}$, Mann–Whitney test. Error bars, s.e.m. **c**, The z-projections of phospho-Akt (green), RFP (red) and Myc (blue) in wing discs upon modulation of Myc levels in the posterior compartment. Scale bar, 100 μm. Note that intensities are not comparable between discs. Representative of ten discs for each. **d**, Wing discs 48 h ACI with clones with low PIP3 (UAS-$pi3kDN$, low PIP3) or high PIP3 (UAS-$pten$ $RNAi$) or loser cells (supercompetition assay) overexpressing a constitutive active form of PI3K (PI3K-CA). Insets show representative clones out of 137, 166 and 268 clones analysed, respectively. White arrows, fragmented clones. Scale bars, 100 μm. **e**, Proportion of fragmented clones; n = number of clones. P values, Fisher's exact tests with WT in WT in blue, and UAS-βgal in tub-$dmyc$ in green; NS, $P > 0.05$; $*P < 0.05$; $**P < 10^{-2}$; $***P < 10^{-3}$; $****P < 10^{-4}$. WT in WT and WT in tub-$dmyc$ come from Fig. 2d.
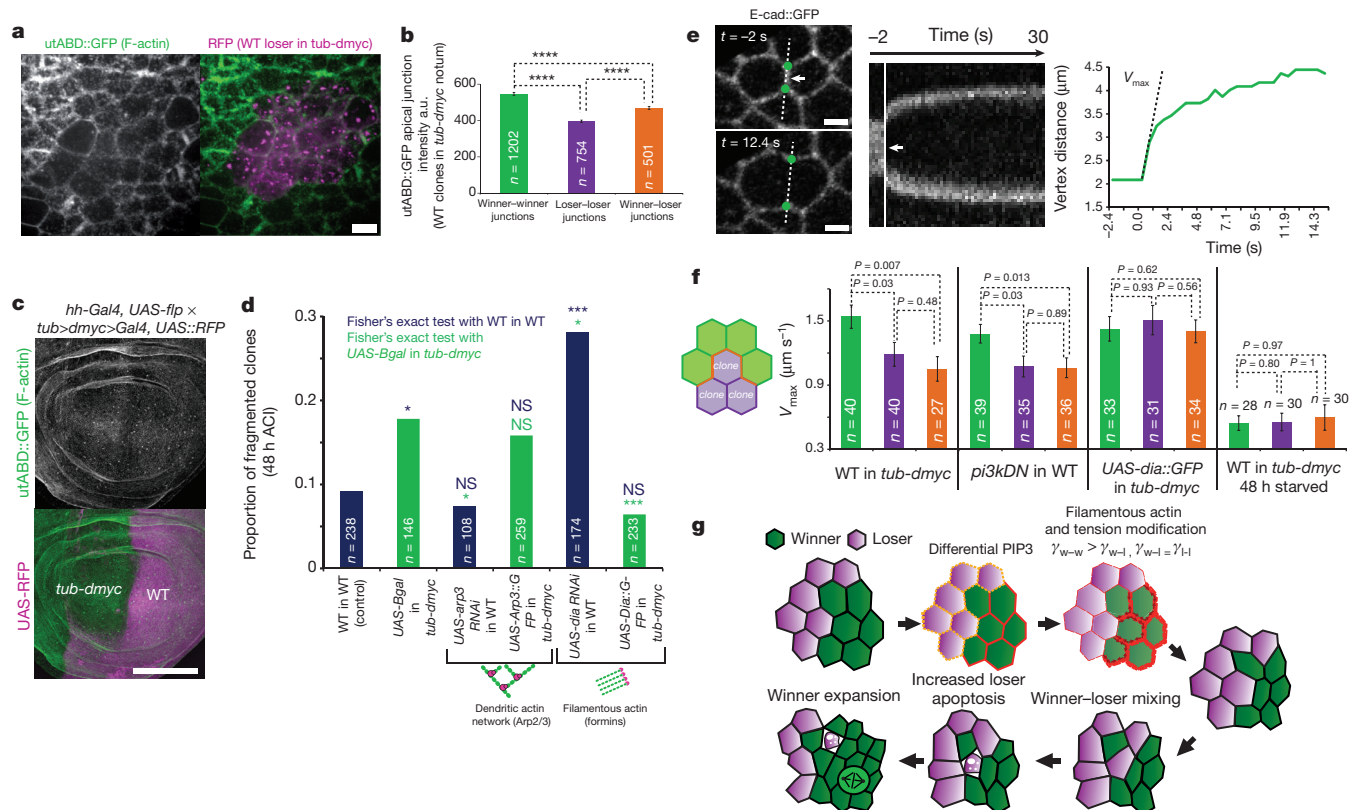
Figure 4 | Filamentous actin and tension modulation are responsible for winner–loser mixing. a, The z-projection around junction plane in the pupal notum 48 h ACI of utrophin actin binding domain fused to GFP (utABD::GFP, F-actin labelling) during supercompetition (purple cells: losers, representative of 42 clones). Scale bar, 5 μm. **b**, Mean junctional utABD::GFP intensity; a.u., arbitrary units; n = number of junctions. Error bars, s.e.m. P values, Mann–Whitney tests; ****$P < 10^{-4}$. **c**, The z-projections of utABD::GFP (green) and RFP (magenta) in a wing disc after removal of one additional copy of *myc* in the posterior compartment (representative of 16 discs). Scale bar, 100 μm. **d**, Proportion of fragmented clones upon perturbation of actin; n = number of clones. Discs are shown in Extended Data Fig. 9b. NS, $P > 0.05$; *$P < 0.05$; **$P < 10^{-2}$; ***$P < 10^{-3}$; ****$P < 10^{-4}$. Fisher's exact tests with

WT in WT in blue, and *UAS-βgal* in *tub-dmyc* in green. WT in WT and WT in *tub-dmyc* come from Fig. 2d. **e**, Junction laser nanoablation in the notum before (top) and 12.4 s after (bottom) ablation. White arrow, ablation point; green points, vertices. Representative of 40 ablation experiments. Right: kymograph showing the evolution of distance between vertices over time and vertex distance plotted over time. The maximum speed of relaxation after ablation (V_{max}) is calculated for each curve and used to assess interfacial tension (see Methods). **f**, Mean V_{max} after junction ablation (green, winner–winner; purple, loser–loser; orange, winner–loser); n = number of junctions; error bars, s.e.m.; P values, Mann–Whitney tests. **g**, Model of the winner–loser mixing process (see Supplementary Discussion). γ, interfacial tension; w, winner; l, loser; red arrows, increased interfacial tension; red lines, F-actin.

disc compartment was sufficient to modify actin levels (Fig. 4c), and F-actin levels were higher in WT clones than $M^{-/+}$ cells (Extended Data Fig. 9a). This prompted us to test the role of actin organization in winner–loser mixing. Downregulating the formin Diaphanous (Dia, a filamentous actin polymerization factor[13]) by RNA interference (RNAi) or by using a hypomorphic mutant was sufficient to obtain a high proportion of fragmented clones (Fig. 4d and Extended Data Fig. 9b, 28%, $P < 10^{-4}$; Extended Data Fig. 9c, 39% fragmented) and to reduce clone compactness over time (Extended Data Fig. 5d, e), whereas over-expressing Dia in loser clones prevented clone splitting (*UAS-dia::GFP*; Fig. 4d and Extended Data Fig. 9b) and compactness reduction (Extended Data Fig. 5d, e). This effect was specific to Dia as modulating Arp2/3 complex (a regulator of dendritic actin network[13]) had no effect on clone splitting (Fig. 4d and Extended Data Fig. 9b). Thus, impaired filamentous actin organization was necessary and sufficient to drive loser–winner mixing. These actin defects were driven by the differences in PIP3 levels between losers and winners (Extended Data Fig. 10). Thus Dia could be an important regulator of competition through its effect on cell mixing. Overexpression of Dia was indeed sufficient to reduce loser clone elimination significantly (Extended Data Fig. 9d) without affecting Hippo/YAP–TAZ pathway[25] (Extended Data Fig. 9e).

Filamentous actin has been associated with tension regulation[13]. We therefore asked whether junction tension was modified in winner and loser junctions. The maximum speed of relaxation of junction after

laser nanoablation (which is proportional to tension[26]) was significantly reduced in loser–loser and winner–loser junctions compared with winner–winner junctions (Fig. 4e, f and Supplementary Video 10). This distribution of tension has been proposed to promote cell mixing[27]. Accordingly, decreasing PIP3 in clones reduced tension both in low-PIP3–low-PIP3 and low-PIP3–normal-PIP3 junctions, whereas overexpressing Dia in loser clones or starvation were both sufficient to abolish differences in tension (Fig. 4f and Supplementary Video 10), in agreement with their effect on winner–loser mixing and the distribution of F-actin. Thus the lower tension at winner–loser and loser–loser junctions is responsible for winner–loser mixing. Altogether, we concluded that the relative PIP3 decrease in losers increases winner–loser mixing through Akt-dependent differential growth and the modulation of tension through F-actin downregulation in winner–loser and loser–loser junctions (Fig. 4g).

Several modes of tissue invasion by cancer cells have been described[28], most of them relying on the departure of the tumour cells from the epithelial layer[29]. This study suggests that some oncogenes may also drive tissue destruction and invasion by inducing ectopic cell intercalation between cancerous and healthy cells, and subsequent healthy cell elimination. *myc*-dependent invasion could be enhanced by other mutations further promoting intercalation (such as *PTEN*). Stiffness is increased in many tumours[30], suggesting that healthy cell–cancer cell mixing by intercalation might be a general process.

1. Walck-Shannon, E. & Hardin, J. Cell intercalation from top to bottom. *Nature Rev. Mol. Cell Biol.* **15,** 34–48 (2014).
2. Morata, G. & Ripoll, P. Minutes: mutants of *Drosophila* autonomously affecting cell division rate. *Dev. Biol.* **42,** 211–221 (1975).
3. de la Cova, C., Abril, M., Bellosta, P., Gallant, P. & Johnston, L. A. *Drosophila* myc regulates organ size by inducing cell competition. *Cell* **117,** 107–116 (2004).
4. Moreno, E. & Basler, K. dMyc transforms cells into super-competitors. *Cell* **117,** 117–129 (2004).
5. Sancho, M. *et al.* Competitive interactions eliminate unfit embryonic stem cells at the onset of differentiation. *Dev. Cell* **26,** 19–30 (2013).
6. Claveria, C., Giovinazzo, G., Sierra, R. & Torres, M. Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* **500,** 39–44 (2013).
7. Levayer, R. & Moreno, E. Mechanisms of cell competition: themes and variations. *J. Cell Biol.* **200,** 689–698 (2013).
8. Tamori, Y. *et al.* Involvement of Lgl and Mahjong/VprBP in cell competition. *PLoS Biol.* **8,** e1000422 (2010).
9. Martins, V. C. *et al.* Cell competition is a tumour suppressor mechanism in the thymus. *Nature* **509,** 465–470 (2014).
10. Moreno, E., Basler, K. & Morata, G. Cells compete for decapentaplegic survival factor to prevent apoptosis in *Drosophila* wing development. *Nature* **416,** 755–759 (2002).
11. Merino, M. M. *et al.* Elimination of unfit cells maintains tissue health and prolongs lifespan. *Cell* **160,** 461–476 (2015).
12. Rhiner, C. *et al.* Flower forms an extracellular code that reveals the fitness of a cell to its neighbors in *Drosophila*. *Dev. Cell* **18,** 985–998 (2010).
13. Lecuit, T., Lenne, P. F. & Munro, E. Force generation, transmission, and integration during cell and tissue morphogenesis. *Annu. Rev. Cell Dev. Biol.* **27,** 157–184 (2011).
14. Simpson, P. & Morata, G. Differential mitotic rates and patterns of growth in compartments in the *Drosophila* wing. *Dev. Biol.* **85,** 299–308 (1981).
15. Landsberg, K. P. *et al.* Increased cell bond tension governs cell sorting at the *Drosophila* anteroposterior compartment boundary. *Curr. Biol.* **19,** 1950–1955 (2009).
16. Simpson, P. Parameters of cell competition in the compartments of the wing disc of *Drosophila*. *Dev. Biol.* **69,** 182–193 (1979).
17. Li, W., Kale, A. & Baker, N. E. Oriented cell division as a response to cell death and cell competition. *Curr. Biol.* **19,** 1821–1826 (2009).
18. Bardet, P. L. *et al.* PTEN controls junction lengthening and stability during cell rearrangement in epithelial tissue. *Dev. Cell* **25,** 534–546 (2013).
19. Andersen, D. S., Colombani, J. & Leopold, P. Coordination of organ growth: principles and outstanding questions from the world of insects. *Trends Cell Biol.* **23,** 336–344 (2013).
20. Britton, J. S., Lockwood, W. K., Li, L., Cohen, S. M. & Edgar, B. A. *Drosophila*'s insulin/PI3-kinase pathway coordinates cellular metabolism with nutritional conditions. *Dev. Cell* **2,** 239–249 (2002).
21. Legoff, L., Rouault, H. & Lecuit, T. A global pattern of mechanical stress polarizes cell divisions and cell shape in the growing *Drosophila* wing disc. *Development* **140,** 4051–4059 (2013).
22. Mao, Y. *et al.* Differential proliferation rates generate patterns of mechanical tension that orient tissue growth. *EMBO J.* **32,** 2790–2803 (2013).
23. Kolsch, V., Charest, P. G. & Firtel, R. A. The regulation of cell motility and chemotaxis by phospholipid signaling. *J. Cell Sci.* **121,** 551–559 (2008).
24. Bosveld, F. *et al.* Mechanical control of morphogenesis by Fat/Dachsous/Four-jointed planar cell polarity pathway. *Science* **336,** 724–727 (2012).
25. Sansores-Garcia, L. *et al.* Modulating F-actin organization induces organ growth by affecting the Hippo pathway. *EMBO J.* **30,** 2325–2335 (2011).
26. Rauzi, M., Verant, P., Lecuit, T. & Lenne, P. F. Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nature Cell Biol.* **10,** 1401–1410 (2008).
27. Brodland, G. W. & Veldhuis, J. H. The mechanics of metastasis: insights from a computational model. *PLoS ONE* **7,** e44281 (2012).
28. Parisi, F. & Vidal, M. Epithelial delamination and migration: lessons from *Drosophila*. *Cell Adhes. Migr.* **5,** 366–372 (2011).
29. Leung, C. T. & Brugge, J. S. Outgrowth of single oncogene-expressing cells from suppressive epithelial environments. *Nature* **482,** 410–413 (2012).
30. Butcher, D. T., Alliston, T. & Weaver, V. M. A tense situation: forcing tumour progression. *Nature Rev. Cancer* **9,** 108–122 (2009).

Author Contributions R.L. and E.M. designed the experiments. R.L. performed and analysed the experiments. B.H. generated *fwe* knockout and *fwe*^*loseA*::*mcherry* knock-in flies. R.L. and E.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M. (eduardo.moreno@izb.unibe.ch).

METHODS

Fly stocks and clone induction. The following stocks were used in this study: *hs-flp22; tub<dmyc<Gal4; UAS-mcd8::GFP* (ref. 4), *hs-flp22; tub<dmyc<Gal4,ubi-Ecad::GFP; UAS-mcd8::RFP* (ubi-Ecad::GFP from ref. 31, stock generated for this study), *UAS-diap1* (third, Bloomington), *UAS-fwe^{loseA}::HA* (on third[12]), *UAS-fwe^{loseA/B}* short hairpin RNAi (third[32]), *hs-flp22; act<y+<Gal4, UAS–GFP* (second, Bloomington), *hs-flp22; act<y+<Gal4, UAS-mcd8::RFP* (second, Bloomington); *hs-flp22; ubi-Cad::GFP, UAS-mRFP; act<y+<Gal4* (a gift from L. Legoff), *UAS-βgal* (second, Bloomington), *UASp-Ecad* (X[33]), *UASt-sqhE20E21* (second[34]), *GMR-gal4, UAS-eiger* (second[35]), *UAS-p35* (second, Bloomington), *zipper^2* (amorphic allele, Bloomington), *zipper^{ebr}* (hypomorphic allele, a gift from O. Baumann), *dpp-gal4* (third, Bloomington), *hs-flp; UAS-fwe^{loseB}::HA; act<cd2<Gal4, UAS–GFP*[12], *act-gal4 switch 255* (second[36]), *FRT40A ubi-nlsGFP* (Bloomington), *FRT40A arm-βgal* (Bloomington), *FRT82B, arm-LacZ* (Bloomington), *FRT82B, RpS3, ubi-nlsGFP* (Minute mutant, Bloomington), *ubi-ptGPH::GFP* (second[20]), *UAS-myc* (third, Bloomington), *UAS-myc RNAi* (second, Bloomington), *hh-gal4,UAS-flp* (recombined third[37], and Bloomington), *UAS-ptenRNAi* (third, Bloomington), *UAS-pi3kDN* (second, Bloomington), *UAS-pi3k^{CAAX}* (pi3kCA, X, Bloomington), *UAS-akt RNAi* (second, Bloomington), *UAS-akt* (second, Bloomington), Δ[*dilps_{1-5}*] (third[38]), *sqh-utABD::GFP* (second or third[39]), *UAS-Arp3 RNAi* (second, TRiP Bloomington), *UAS-Arp3::GFP* (second, Bloomington), *UAS-dia RNAi* (second, TRiP Bloomington), *UAS-dia::GFP* (third[40]), *FRT40A dia^5* (Bloomington, hypomorphic allele[40]), *endo-Ecad::GFP* (knock-in[41]), *sqh-Sqh::GFP* (second[42]), *dachs-Dachs::GFP* (third[24]), *expanded-LacZ* (second, Bloomington), *fwe^{loseA}::mcherry* KI (generated for this study), *eng-Gal4, UAS–RFP* (second, Bloomington).

We used both male and female larvae and pupae for experiments. All heat shocks were performed in a 37 °C waterbath using glass tubes. For notum live imaging, clones were generated using the *hs-flp22; tub<dmyc<Gal4, ubi-Ecad::GFP; UAS-mcd8::RFP* line with a 35 min heat shock or for 4 h (1 h four times, with 1 h rest in-between) for measuring the shape evolution of *tub-dmyc* patches surrounded by WT (Extended Data Fig. 5d, e), 12 min for *hs-flp; act<y+<Gal4, UAS-mcd8::RFP* (crossed with ubi-Cad::GFP for WT in WT control) and *hs-flp; ubi-Cad::GFP, UAS-mRFP; act<y+<Gal4* (crossed with *UAS-fwe^{loseA}, UAS-pi3kDN, UAS-pten RNAi, UAS-dia RNAi*) lines. For the supercompetition assay in the wing disc (WT clones in *tub-dmyc*), larvae were heat shocked for 20 min for videos, or for 14 min (Extended Data Fig. 3a–c) or 12 min (Figs 2–4 and Extended Data Figs 4c 6h–k, 9b, d and 10e, f). For WT clone in WT background (*hs-flp22; act<y+<Gal4, UAS–GFP*), we used 5 min 30 s heat shock for Extended Data Fig. 3a–c and 5 min for all other experiments. *UAS-fwe^{loseB}::HA; Act<cd2<Gal4* clones were induced with a 8 min heat shock (Extended Data Fig. 2). For *FRT40A ubi-nlsGFP/FRT40 arm-βgal* and *FRT40A dia5/FRT40 ubi–GFP* clones, larvae were heat shocked for 5 min. We used 30 min heat shocks to generate WT clones in $M^{-/+}$ background. Larvae were collected and dissected 24, 48 or 72 h after heat shock at the L3 wandering stage. For starvation experiments, larvae were collected 24 h or 48 h ACI, washed in distilled water and left on humidified paper without food for 24 h (48 h for imaging pupae). Pupae were dissected and mounted as indicated in ref. 43, 48 or 72 h ACI and 18–20 h after pupae formation.

Activation of progesterone-sensitive gal4 (gal4 switch) was done by using fly food mixed with RU486 at 1 µg ml^{-1} or 50 µg ml^{-1}. Full development occurred in the hormone-containing food (egg laying and larval development). *Act<cd2<Gal4* clones were induced with an 8 min heat shock.

Generation of *Flower^{loseA}::mcherry* knock-in. The Flower knock-in fly was made by genomic engineering[41]. The genomic engineering in ref. 41 is a two-step process consisting of ends-out gene targeting followed by phage integrase phi31-mediated DNA integration. A founder knockout line was established with a genomic deletion of the *flower* locus at position 3L: 15816737–15810028. The knock-in construct was composed of the deleted *flower* locus with a mCherry fusion after exon 5 (see Extended Data Fig. 2a, specific for *loseA* isoform). The knock-in construct was done by site-directed mutagenesis to remove the stop codon and add a restriction site, used to insert mCherry after exon 5. The knockout of *flower* and the knock-in *Flower^{loseA}::mcherry* were tested by PCR and sequencing.

The vectors used for generating the *FlowerloseA::mcherry* were as follows: pGX-attP, knockout vector; pGEM-T, used for the site-directed mutagenesis; pGEattB^{GMR}, knock-in vector.

Primers used were as follows: 5′ arm Fwe forward, AAGCGGCCGCAG CAGCAACAACAGCAGCAACG; 5′ arm Fwe reverse, AAGCGGCCGCACCGT TCAATATGCAGGCGGC; 3′ arm Fwe forward, GGAGATCTGGATGAT TCCTGAGCTGCGGTAT; 3′ arm Fwe reverse, AACTGCAGATGGGGACACC TAAAGGCAGGCACC; Fwe KI forward, ATAGATCTATAAAAGCTTTCAATGT ACACAAATTTG; Fwe KI reverse, ATGGTACCAAAAAGCATGCCCCACAAT AGTTAC; Mutagenesis primer forward, TCTGGCGCCCCACAGGGCCGG CCATTCCTAGCCGCGCC; Mutagenesis primer reverse, TGGCGCGGCTA GGAATGGCCGGCCCTGTGGGGCGCC; mCherry forward, TAGGGCCGG CCATGGTGTCCAAGGGCGAAG; mCherry reverse, ATGGCCGGCCCTTAT TTATACAGCTCGTCCATGC.

Immunohistochemistry and image acquisition. Immunostainings of the wing discs were performed with standard formaldehyde fixation and permeabilization/washes in PBT 0.4% Triton. The following antibodies/markers were used: rat anti-E-cad (DCAD2 concentrate, Developmental Studies Hybridoma Bank, 1/50); chicken anti-GFP (1/200, Abcam); guinea-pig and rabbit anti-Dia (1/50, a gift from J. Grosshans[44]); rabbit anti-phospho-Akt (1/50, Cell Signalling); guinea-pig anti-Myc (1/500, a gift from G. Morata); phalloidin Alexa 546 (1/50, Life Technologies); mouse anti-Ptc (1/50, Developmental Studies Hybridoma Bank); rabbit anti-β-gal (1/200, Cappel); rat anti-HA (1/100, Roche). Dia and phospho-Akt stainings were amplified using biotinylated secondary antibodies and streptavidin coupled to Cy3 or Cy5. TUNEL stainings used the Roche kit.

Dissected wing discs were mounted in Vectashield (Vectorlab) and imaged on a Leica confocal SP2 using a ×63 water immersion objective numerical aperture 1.3 or a Leica confocal SP8 using a ×63 oil objective with numerical aperture 1.4. Images shown are maximal z-projections containing the adherens junction plane.

Adult eye pictures were taken on a Leica MZFLIII dissecting scope with a Nikon DXM1200 colour camera at the same magnification. Eye area was measured in Fiji[45].

Pupal notum and wing disc live imaging, death probability calculation. Pupae were collected 48 or 72 h ACI and dissected 18–20 h after pupae formation. Pupae were prepared as indicated in ref. 43 and imaged on a confocal spinning disc microscope (Till Photonics) with a ×40 oil objective (numerical aperture 1.35) or a point scanning confocal microscope Leica SP8 with a ×63 objective (numerical aperture 1.4). Large-field views of the tissue were obtained by tile imaging (6–12 tiled positions). The z-stacks (1 µm per slice) were recorded every 5 min using autofocus at every cycle and every tiled position using E-cad::GFP signal. Videos were recorded in the nota close to the scutellum region in the vicinity of the aDT and pDT macrochaetes. Videos shown are cropped from larger fields of view after maximum projections, correction for bleach (using Fiji) and correction for tissue drift (Stack reg plugin, Fiji).

Every cell delamination event in the clones was tracked over 10 h. We excluded cells in the midline where spontaneous delaminations were occurring[46]. The proportion of apical perimeter shared with winners (sum of the winner–loser junction lengths over the total apical perimeter) was measured 1 h before delamination at the junction plane (E-cad::GFP signal or actin belt visualized with utABD::GFP) using imageJ. We then measured the proportion of apical perimeter shared with winners for all the cell in the clones at time 0 in every video (excluding cells in the midline) manually or by using CellPacking analyser for skeletonization[47] and a home-made macro (Igor Pro Software, Wavemetrics). Probabilities of death were then obtained by dividing the number of delaminating cells by the total number of loser cells in each shared perimeter category. Note that similar correlations were obtained by using the number of junctions shared with winners divided by total number of junctions (not shown).

Ex vivo culture of wing-discs used clone8 media, as indicated in ref. 21, using a point scanning confocal microscope Leica SP8 with a ×63 objective (numerical aperture 1.4), 36 h ACI (WT in *tub-dmyc*). Discs showing rapid drop-down of cell division were excluded from the analysis. Removal of the signal from the peripodal cells and selection of the signal from the junction plane were performed by using a home-made Matlab macro for selective plane projection (inspired by ref. 21). For every x–y pixel, the z-plane with the maximal E-cad::GFP intensity (calculated by summing pixel intensity on a 50 pixel × 50 pixel square at every plane) was kept and used to retrieve RFP signal in the same plane. The measurement of probability of death was performed as in the notum. Imaging of *fwe^{loseA}::mcherry* KI was also performed in *ex vivo* wing discs using the same projection procedure.

Junction remodelling was manually counted at the interface of the clones. Each event leading to the disappearance of a junction between two RFP-positive cells was counted. Remodelling events were only counted if the new topology was maintained until the end of the video (10 h; Fig. 2f and Extended Data Fig. 5c). The total number of remodelling events was then divided by the total number of junctions analysed. For Extended Data Fig. 5a, we counted every remodelling event occurring over 10 h for single junctions, and calculated the proportion of junctions undergoing a single remodelling event and the probability of undergoing an additional remodelling event (after a first remodelling event). All winner–winner junctions tracked shared one vertex with a loser cell, whereas loser–loser junctions tracked also shared one vertex with a winner cell. All winner–loser junctions were tracked.

Image quantification. Clone fragmentation, clone size and clone compactness. Clones were counted as fragmented when GFP-positive cells were separated by a single GFP-negative cell in the apical area (using E-cad or phalloidin staining) 48 h

ACI in the wing pouch. Discs showing too high a clone density were excluded (>20 clones per wing pouch). This technique accurately evaluated the number of clones that had recently split. The probability of obtaining false positive results was expected to be low because the probability of observing two independent clones separated by a single cell required the initial recombination to occur in two cells separated by a single cell, as well as the absence of division of the cell in-between for the subsequent 48 h (on average, cells should divide more than three times[48]). Accordingly, we did not observe lower fragmentation when it was evaluated using twin clones (Extended Data Fig. 4a, b) where 2×β-gal-positive clones were counted as fragmented when cells were separated by single negative cell, and when the sister clone (2×GFP) formed a single continuous group. However, we slightly underestimated the total number of fragmented clones because the proportion of all the split 2×β-gal clones was close to 12% (including groups separated by more than one cell) (Extended Data Fig. 4a, b).

Clone size was measured on 9-μm-wide maximum projection containing the apical junction plane in the wing pouch. GFP-positive patches were automatically recognized and segmented with a homemade Fiji macro, and used to measure the average clone area, the density of clones ((number of clones/area of the wing pouch) × 10,000 = number of clones per 10,000 μm²) and the total surface of the wing pouch covered by GFP-positive cells (total clone area/area of the wing pouch).

Clone compactness is defined as $C = 4\pi \times (\text{area/perimeter}^2)$. Clone compactness was measured for every clone in the notum by drawing their contour using Fiji at t_0 and 10 h later. The fold change was calculated as $(C_{10h} - C_{0h})/C_{0h}$.

Death induction at compartment boundary. Measurement of death induction at the antero-posterior compartment boundary was done by counting the number of TUNEL-positive cells in the wing pouch in the posterior compartment (control compartment) and in an ROI in the anterior compartment (a band having a width of three cells along the antero-posterior compartment boundary). The compartment boundary was detected with Patched staining. The z-projections of the three different genotypes (*dpp-gal4*;*UASfweloseA::HA*; *zip²/zipᵉᵇʳ*; and *zip²/zipᵉᵇʳ*; *dpp-gal4/UASfweloseA::HA*) were randomized by assigning random file names, and TUNEL cell counting was performed blindly. The expected random distribution of dead cells in the ROI was calculated by taking the average ratio of surface of the ROI over the total surface of the wing pouch for ten wing discs. The *zip²/zipᵉᵇʳ* larvae were sorted by using a fluorescent balancer chromosome (*Cyo, act–GFP*, Bloomington).

Intensity measurements. The intensity measurements for *fweloseA::mcherry* were done in *ex vivo* cultured wing disc on selective plane projections (as described previously). For each cell in the loser clones, we measured the membrane intensity using Fiji (with a width line of six pixels) divided by the average membrane intensity of all the cells measured in the same disc, and measured the proportion of perimeter shared with winner cells.

PIP3 intensity measurements were performed on maximum projections of cell apical area (4 μm) in living notum expressing ubi-tGPH::GFP. Cytoplasmic signal was removed using two subsequent background substractions on Fiji (Rolling Ball radius 200 pixels and 5 pixels). Junction signal was the mean intensity on a line of width 6 pixels (1 pixel = 0.148 μm). Measurements of utABD::GFP junction intensity were performed similarly on a line of width 6 pixels after a single background subtraction (rolling ball radius 200 pixels).

Measurements of line intensity profile in the wing discs (Extended Data Fig. 10) were performed by using a line of width 100 pixels in the dorsal compartment parallel to the dorso-ventral compartment boundary (1 pixel = 0.267 μm) on maximal z-projections using Fiji. Intensity profiles were measured for Patched (Ptc) and utABD::GFP or Dia. Position 0 was determined by detecting the minimum of the derivative of Ptc intensity profile (calculated on Igor Pro Software after smoothing of the profile, 300 pixels averaging window), which corresponded to the boundary between the anterior and the posterior compartments. The distance from the most anterior side of the profile to the antero-posterior boundary was normalized to 1. Each intensity profile was divided by its mean. The average intensity profile was then calculated by averaging the normalized intensity values obtained for each disc at the same relative distance to the antero-posterior boundary.

FRAP. The utABD::GFP has been previously used to assess actin dynamics[39,49]. FRAP experiments with utABD::GFP were performed in *ex vivo* cultured wing disc in dimethylsulfoxide (DMSO; 0.2%, control) and after treatment with Jasplakinolide (2 μM, Life Technologies) using a confocal spinning disc microscope (Till Photonics) and a ×60 oil objective (numerical aperture 1.35). Bleaching was done using 100% of the power of a 488 solid-state laser in diffraction-limited ROI after acquiring three time points (one frame per 0.5 s). Recovery was recorded on 30 s. Intensity recoveries were obtained by measuring the mean intensity in an ROI of 15 pixels × 15 pixels in Fiji (1 pixel = 0.10 μm) containing the bleached region. Each curve was normalized by the intensity profile of a neighbouring

control region (20 pixels × 20 pixels) to correct for bleaching due to imaging. For each normalized curve, we subtracted the intensity at t_0 (post-bleach) and then divided by the average intensity of the three first points (pre-bleach). The averages of all the recovery curves were then shown. FRAP in the notum was performed similarly 48 h ACI (WT losers in *tub-dmyc*). Characteristic times of recovery were calculated by fitting the normalized curves with Igor Pro software using the equation $I(t) = A - B \exp(-t/\tau)$, where t is the time (in seconds), τ is the characteristic time of recovery, A the mobile fraction and $A - B$ the initial intensity after bleaching. For comparison between winner–winner and winner–loser junctions in the same cell, we performed simultaneous bleaching and recovery recording in two identical ROIs (one in a winner–winner junction, one in a winner–loser junction of the same cell). Each recovery curve was normalized and fitted as mentioned above. We then calculated the fold change of characteristic time of recovery for each cell: $(\tau_{w-w} - \tau_{w-l})/\tau_{w-w}$.

The *endo-Ecad::GFP* FRAP experiments were performed similarly using a Leica SP8 point scanning confocal microscope (×63 oil immersion objective, numerical aperture 1.4), using an argon laser for imaging and bleaching (488 nm). Recovery was assessed over 200 s, one frame per second. The average recovery curves were obtained in the same way as for the *utABD::GFP* FRAP experiments.

Laser ablation. Junction laser ablations were performed in the notum using a two-photon infrared laser on a two-photon microscope (two-photon FluoView 1000 Olympus, Center for Microscopy and Image Analysis, University of Zurich) using a ×25 water objective (numerical aperture 1.05). Imaging and ablation were performed with 950 nm wavelength, 1.3% power for imaging, 30% power and 300 ms exposition for the ablation scanning along a line perpendicular to the junction. Relaxation of vertices (visualized with E-cad::GFP) was tracked for at least 10 s (one frame per 0.594 s, one frame per 0.891 s for Fig. 4f 'starved'). Positions of vertices were then tracked using CellTrack software[50] and the evolution of vertex distance over time was fitted with an exponential function $(A_0 + A_1\exp(A_2t))$ on the first ten time points after ablation using Igor software. The values of V_{max} were the derivatives of the exponential fit at t_0.

Statistics. All the error bars shown in the figures are s.e.m. or 95% confidence intervals for probabilities of death. Statistical tests were Fisher's exact tests (two-sided) for comparison of proportions or Mann–Whitney non-parametric tests for all other experiments, except in Extended Data Fig. 8f where we used a one-sample *t*-test with 0 as reference value (normality tested with a Shapiro–Wilk test). Difference in variance was always below 2. No statistical methods were used to set sample size. Experiments were not randomized, and we did not analyse experiments blindly, except for the result of Fig. 1h. Every experiment was done at least three independent times. The same control was used for all the quantifications of clone splitting because it was based on three independent experiments and gave reproducible values (for example compare Fig. 2d, WT in WT, and Extended Data Fig. 4a, b).

31. Oda, H. & Tsukita, S. Real-time imaging of cell-cell adherens junctions reveals that *Drosophila* mesoderm invagination begins with two phases of apical constriction of cells. *J. Cell Sci.* **114,** 493–501 (2001).

32. Merino, M. M., Rhiner, C., Portela, M. & Moreno, E. "Fitness fingerprints" mediate physiological culling of unwanted neurons in *Drosophila. Curr Biol.* **23,** 1300–1309 (2013).

33. Pacquelet, A. & Rorth, P. Regulatory mechanisms required for DE-cadherin function in cell migration and other types of adhesion. *J. Cell Biol.* **170,** 803–812 (2005).

34. Corrigall, D., Walther, R. F., Rodriguez, L., Fichelson, P. & Pichaud, F. Hedgehog signaling is a principal inducer of myosin-II-driven cell ingression in *Drosophila* epithelia. *Dev. Cell* **13,** 730–742 (2007).

35. Moreno, E., Yan, M. & Basler, K. Evolution of TNF signaling mechanisms: JNK-dependent apoptosis triggered by Eiger, the *Drosophila* homolog of the TNF superfamily. *Curr Biol.* **12,** 1263–1268 (2002).

36. Ford, D. *et al.* Alteration of *Drosophila* life span using conditional, tissue-specific expression of transgenes triggered by doxycycline or RU486/Mifepristone. *Exp. Gerontol.* **42,** 483–497 (2007).

37. Tanimoto, H., Itoh, S., ten Dijke, P. & Tabata, T. Hedgehog creates a gradient of DPP activity in *Drosophila* wing imaginal discs. *Mol. Cell* **5,** 59–71 (2000).

38. Zhang, H. *et al.* Deletion of *Drosophila* insulin-like peptides causes growth defects and metabolic abnormalities. *Proc. Natl Acad. Sci. USA* **106,** 19617–19622 (2009).

39. Rauzi, M., Lenne, P. F. & Lecuit, T. Planar polarized actomyosin contractile flows control epithelial junction remodelling. *Nature* **468,** 1110–1114 (2010).

40. Homem, C. C. & Peifer, M. Diaphanous regulates myosin and adherens junctions to control cell contractility and protrusive behavior during morphogenesis. *Development* **135,** 1005–1018 (2008).

41. Huang, J., Zhou, W., Dong, W., Watson, A. M. & Hong, Y. Directed, efficient, and versatile modifications of the *Drosophila* genome by genomic engineering. *Proc. Natl Acad. Sci. USA* **106,** 8284–8289 (2009).

42. Royou, A., Field, C., Sisson, J. C., Sullivan, W. & Karess, R. Reassessing the role and dynamics of nonmuscle myosin II during furrow formation in early *Drosophila* embryos. *Mol. Biol. Cell* **15,** 838–850 (2004).

43. Bellaiche, Y., Gho, M., Kaltschmidt, J. A., Brand, A. H. & Schweisguth, F. Frizzled regulates localization of cell-fate determinants and mitotic spindle rotation during asymmetric cell division. *Nature Cell Biol.* **3,** 50–57 (2001).
44. Grosshans, J. *et al.* RhoGEF2 and the formin Dia control the formation of the furrow canal by directed actin assembly during *Drosophila* cellularisation. *Development* **132,** 1009–1020 (2005).
45. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (2012).
46. Marinari, E. *et al.* Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature* **484,** 542–545 (2012).
47. Aigouy, B. *et al.* Cell flow reorients the axis of planar polarity in the wing epithelium of *Drosophila*. *Cell* **142,** 773–786 (2010).
48. Martin, F. A., Herrera, S. C. & Morata, G. Cell competition, growth and size control in the *Drosophila* wing imaginal disc. *Development* **136,** 3747–3756 (2009).
49. Burkel, B. M., von Dassow, G. & Bement, W. M. Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin. *Cell Motil. Cytoskeleton* **64,** 822–832 (2007).
50. Sacan, A., Ferhatosmanoglu, H. & Coskun, H. CellTrack: an open-source software for cell tracking and motility analysis. *Bioinformatics* **24,** 1647–1649 (2008).
51. Igaki, T. *et al.* Eiger, a TNF superfamily ligand that triggers the *Drosophila* JNK pathway. *EMBO J.* **21,** 3009–3018 (2002).
52. Zhang, C. *et al.* An intergenic regulatory region mediates *Drosophila* Myc-induced apoptosis and blocks tissue hyperplasia. *Oncogene* **34,** 2385–2397 (2014).
53. Bubb, M. R., Spector, I., Beyer, B. B. & Fosen, K. M. Effects of jasplakinolide on the kinetics of actin polymerization. An explanation for certain in vivo observations. *J. Biol. Chem.* **275,** 5163–5170 (2000).
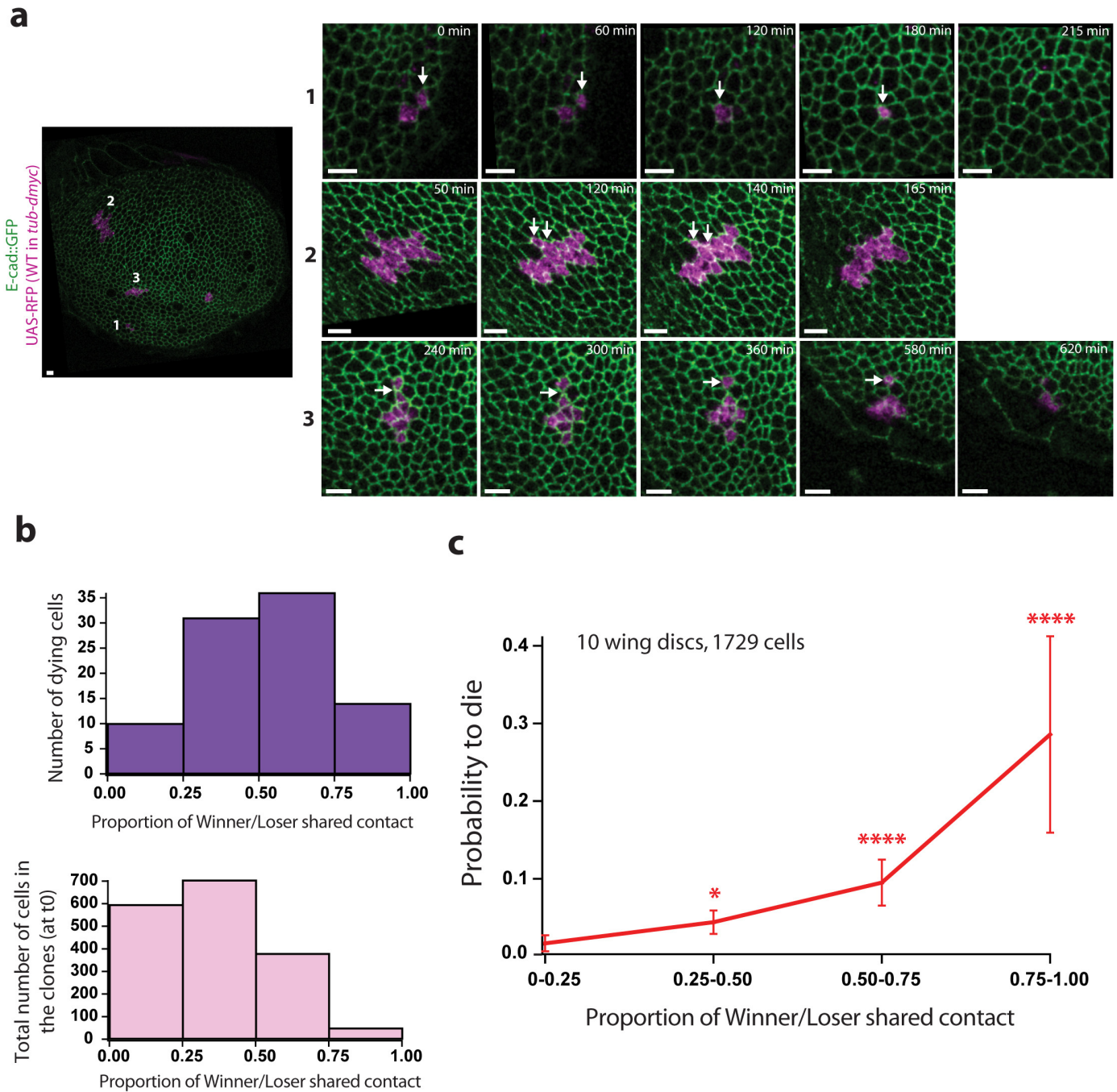
Extended Data Figure 1 | Loser elimination also correlates with the shared perimeter with winners in the wing disc. **a**, Left: selective plane *z*-projection of an *ex vivo* cultured wing disc expressing ubi-Ecad::GFP with loser clones (RFP, purple, WT in *tub-myc*) 36 h ACI, representative of ten videos. Numerals 1, 2, 3 correspond to the clones shown on the right. Right: snapshots showing the delamination of loser cells and one event of clone splitting preceding cell elimination (3). Scale bars, 5 μm. **b**, Distribution of the proportion of junctional perimeter of loser cells shared with winners in eliminated cells 1 h before delamination (top) (0 = cell in the centre of the clone, 1 = isolated cell surrounded by winners) and in all the loser cells at t_0 (bottom). **c**, Probability of loser cell elimination for a given surface of contact shared with winners in WT loser cells in *tub-dmyc*. Statistical tests are Fisher's exact tests performed with the point 0–0.25 (*$P < 0.05$; ****$P < 10^{-4}$). Error bars, 95% confidence interval.
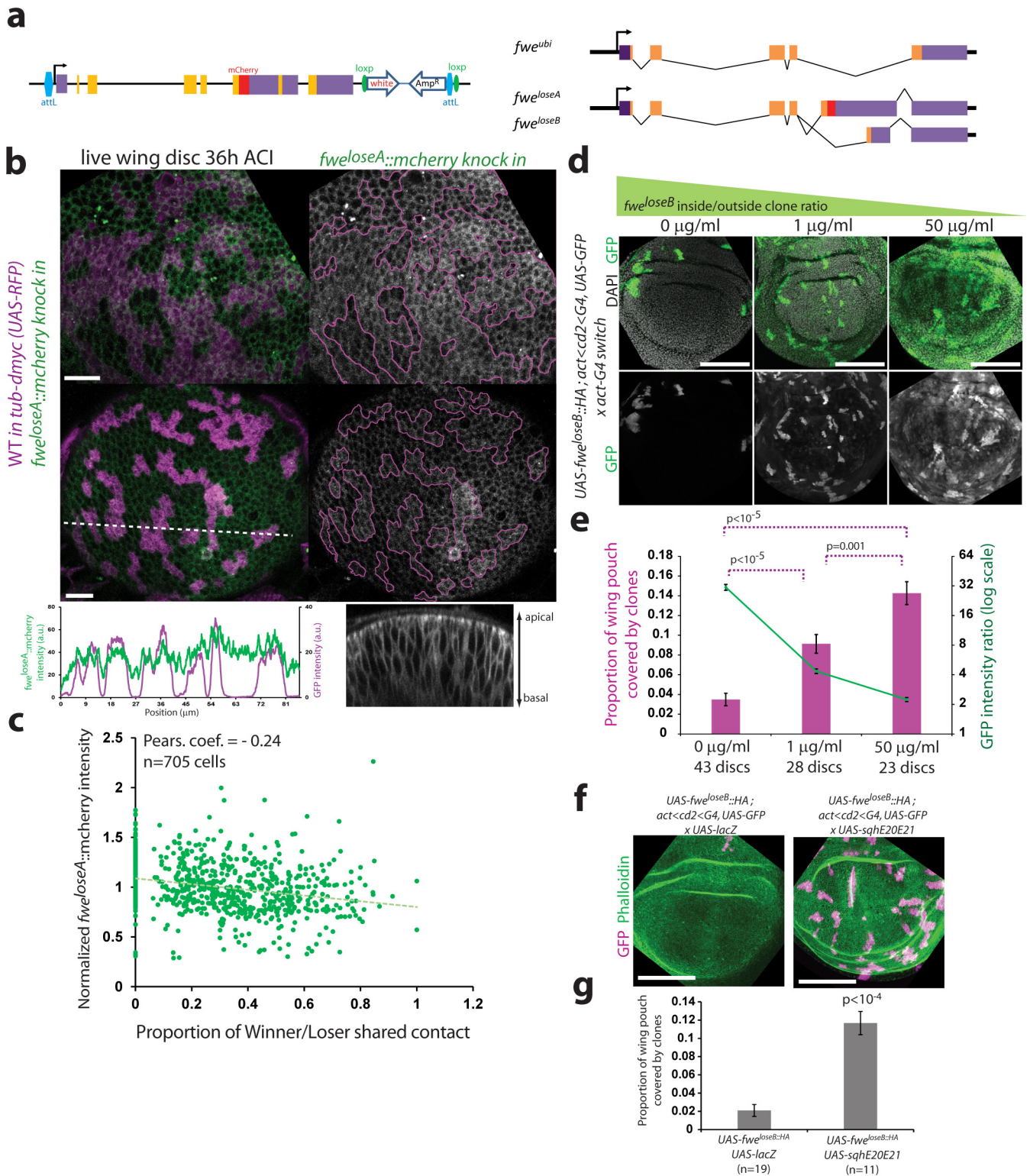
Extended Data Figure 2 | Contact-dependent death is triggered downstream of *flower*. The transmembrane protein Flower is a central regulator of competition[12]. The *fwe^lose* isoforms (*loseA* and *loseB*) are induced downstream of several competition contexts and their expression is necessary for loser elimination and sufficient to drive cell elimination when contacting WT cells[12]. The contact-dependent communication could occur upstream of *fwe* (for instance by modifying the levels of induction of *fwe^lose*) or downstream of *fwe* induction. Several pieces of evidence indicate that it occurs downstream of *fwe* induction. First, cell death also correl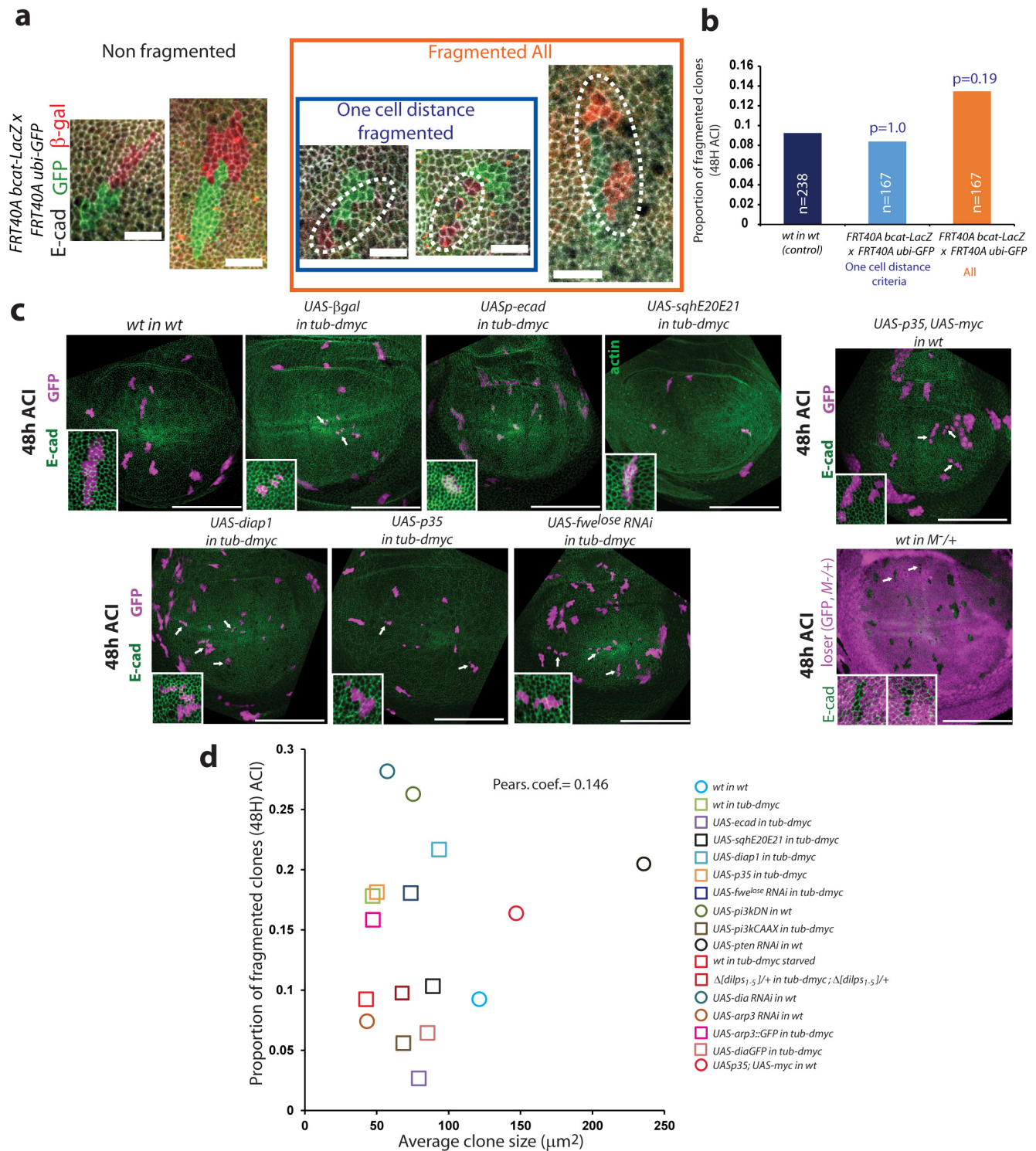ated with shared apical perimeter in clones homogenously expressing *fwe^loseA* (Fig. 1c, red curve). Second, using a knock-in fusion *fwe^loseA::mcherry* (Extended Data Fig. 2a), we could show that *fwe^lose* induction did not correlate with the surface of contact shared with winners (Extended Data Fig. 2b, c) as previously suggested by *in situ* experiments for *fwe^lose* (ref. 12). Finally, the probability of elimination of clones overexpressing *fwe^lose* is proportional to the relative differences in *fwe^lose* levels inside and outside the clones (Extended Data Fig. 2d, e). Altogether, this suggested a model where cells can compute the relative differences of *fwe^lose* levels with all their neighbours through an unknown molecular mechanism. **a**, Schematic of the modified *fwe* locus (left) and the resulting messenger RNA of the three isoforms (right). Orange rectangles are exons. The 5′ and 3′ untranslated regions are shown in purple. Exon 5 is specific to each isoform. The red box shows the localization of the mCherry tag at the end of the exon 5 of *fwe^loseA*. Note that the vector backbone was conserved in the *knock-in* line (white, AmpR). **b**, Two examples of selective plane *z*-projection of *ex vivo* cultured wing discs expressing *fwe^loseA::mcherry KI* in WT clone in *tub-dmyc* background (purple) 36 h ACI representative of 12 discs. The clone contour is shown in purple (right). Scale bars, 10 μm. The intensity profile of the white dotted line is shown below. Bottom right, a lateral view of *fwe^loseA::mcherry* and its accumulation in the apico-lateral region. **c**, Scatter plot of *fwe^loseA::mcherry* membrane intensity in loser cells in wing disc (WT in *tub-myc*, *y* axis) against the proportion of perimeter shared with winner cells (*x* axis). One dot represents one cell. Pearson correlation coefficient = −0.24. **d**, *UAS-fwe^loseB::HA* clones (GFP) in wing discs 72 h ACI with different concentrations of RU486 in the food media. The expression of *fwe^loseB* in clones is the sum of act-G4 flip out driver (constant) and the hormone-sensitive Gal4 (Gal4 switch, expression proportional to the RU486 concentration) while *fwe^loseB* is only driven by the Gal4 switch outside the clone. GFP panels were acquired with the same parameters and are shown with the same contrasts. From left to right, discs representative of 43, 28 and 23 discs. Scale bars, 100 μm. **e**, Average proportion of the wing pouch surface covered by clones (left, purple histogram) and average GFP intensity ratio inside/outside clones (right, green curve, log scale) 72 h ACI. Error bars, s.e.m. Statistical tests are Mann–Whitney tests performed for wing pouch coverage. **f**, *UAS-fwe^loseB::HA* clones (GFP) in wing discs 72 h ACI in a control (left, *UAS-lacZ*, representative of 19 discs) or upon overexpression of active MRLC (*UAS-sqhE20E21*, right, representative of 11 discs). Scale bars, 100 μm. **g**, Average proportion of the wing pouch surface covered by clones 72 h ACI. Error bars, s.e.m. Statistical test is a Mann–Whitney test.
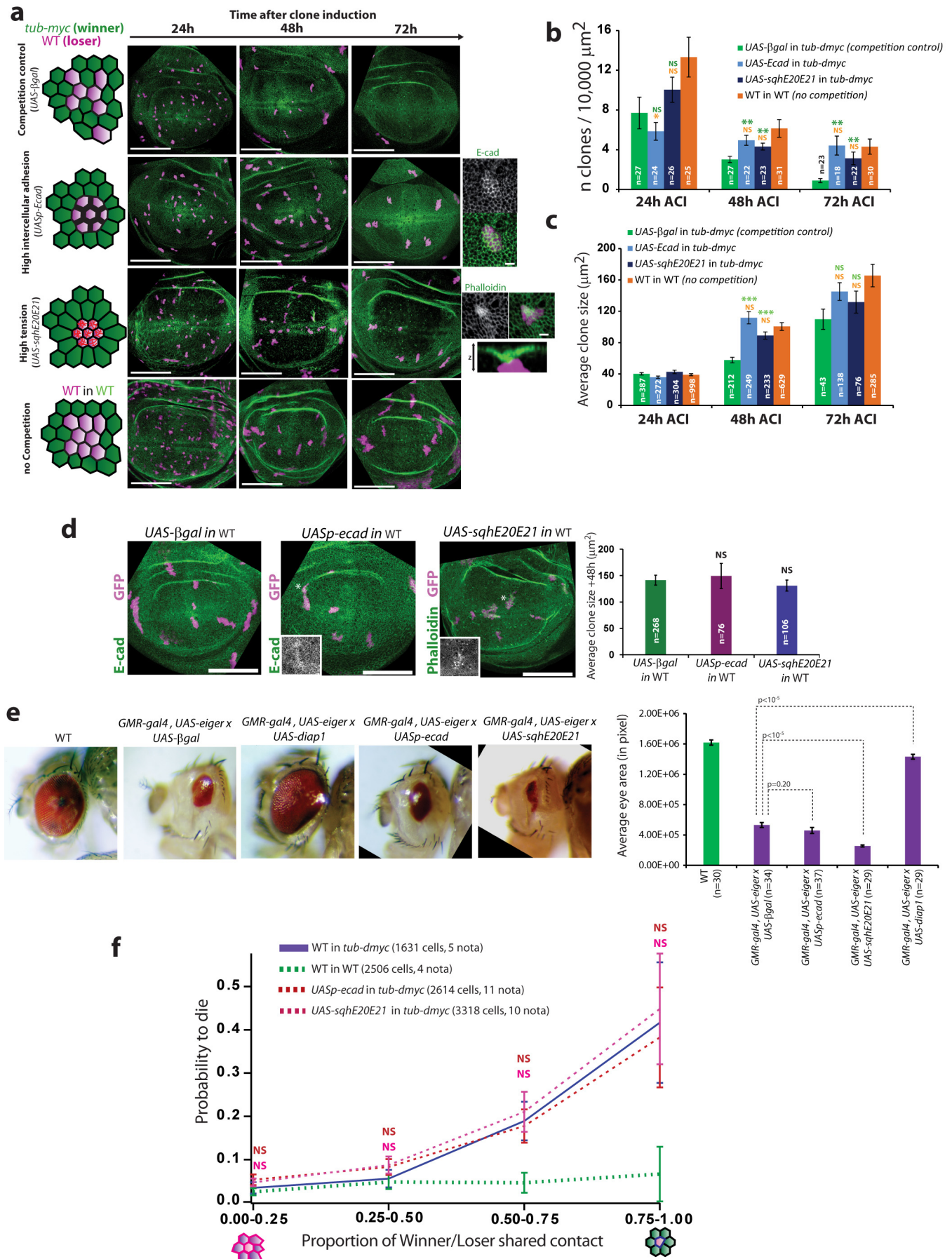
Extended Data Figure 3 | E-cad and active MRLC rescue losers only through the change of winner–loser surface of contact. a, Supercompetition assay in the wing disc 24 h, 48 h and 72 h ACI (purple: loser cells) in normal competition (loser cells overexpressing β-gal), upon limitation of cell mixing (*UAS-ecad* and *UAS-sqhE20E21*, a constitutively active MRLC) and in WT clones in WT background (no competition). Left: schematic of the expected effect on clone shape; black line thickness is the strength of cell–cell adhesion, red lines show actomyosin network. Middle: example of wing discs at different time ACI; scale bars = 100 μm. See Extended Data Fig. 3b for the number of discs scanned for each condition. Right: close-up views of clones overexpressing E-cad (top) and active MRLC (bottom). Active MRLC induces apical actin accumulation and partial apical constriction. Scale bars, 5 μm. **b, c**, Density of loser clones (**b**) and averaged loser clone size (**c**) at 24, 48 and 72 h ACI in the wing pouch. Error bars, s.e.m.; *n* is the number of wing discs (**b**) or the number of clones (**c**). Statistical tests are Mann–Whitney tests performed with the control competition (in green, *UAS-βgal* in *tub-dmyc*) or with control without competition (in orange, WT in WT). NS, $P > 0.05$; *$P < 0.05$; **$P < 10^{-2}$; ***$P < 10^{-3}$. Note that we did not find significant differences of clone size at 72 h as the few clones remaining for the control competition were at the periphery of the wing disc where competition is less effective. **d**, E-cad and active MRLC do not have a cell autonomous effect on growth. Left: wing discs showing WT clones (*act<y<gal4, UAS-mcd8::GFP*, purple) overexpressing *β-gal*, *E-cad* or *sqhE20E21* (active MRLC) 48 h ACI, representative of 56, 13 and 17 discs, respectively. The insets show the clones marked with an asterisk. Scale bars, 100 μm. Right: average clone surface, *n* = number of clones, error bars are s.e.m. NS, $P > 0.05$, Mann–Whitney tests. **e**, E-cad and active MRLC do not prevent genetically induced apoptosis. Left: adult eye of a WT fly (*oregonR*), and flies with abnormal eye morphology due to induction of JNK-dependent death in the eyes (eye-specific gal4, GMR-gal4 and UAS-eiger, the fly orthologue of TNF[35,51]) expressing *β-gal* (control), *diap1* (apoptosis inhibition), *E-cad* or *sqhE20E21* (active MRLC), representative of 30, 34, 37, 29 and 29 adults, respectively. Right: averaged eye surface in pixels (a.u., arbitrary units); *n* = number of flies; error bars, s.e.m. *P* values, Mann–Whitney tests. **f**, E-cad and active MRLC do not modify the probability of loser death for a given surface of contact with winners. Probability of loser cell elimination in the pupal notum for a given surface of contact shared with winners in *myc*-dependent competition (purple, from Fig. 1d), in WT cells in WT background (control, dotted green, from Fig. 1d), or in losers overexpressing E-cad (dotted red) or active MRLC (*sqh-E20E21*, dotted pink). Statistical tests are Fisher's exact tests performed with *myc*-dependent competition (purple) (NS, $P > 0.05$). Error bars, 95% confidence intervals.

a Non fragmented / Fragmented All / One cell distance fragmented

b

c

wt in wt | *UAS-βgal in tub-dmyc* | *UASp-ecad in tub-dmyc* | *UAS-sqhE20E21 in tub-dmyc* | *UAS-p35, UAS-myc in wt*

*UAS-diap1 in tub-dmyc* | *UAS-p35 in tub-dmyc* | *UAS-fwe^lose^ RNAi in tub-dmyc* | *wt in M⁻/+*

d

Extended Data Figure 4 | Clone fragmentation does not correlate with clone size. a, Twin clones 48 h ACI marked with two copies of GFP (green) and absence of β-gal or two copies of β-gal (red) and absence of GFP (*FRT40A ubi-nlsGFP/FRT40A bcat-βgal*). Left: non-fragmented clones. Middle and right: fragmented clones (the GFP sibling clone is used as a reference) with clone cells separated by a single cell (middle) or more than one cell (right). **b**, Proportion of fragmented clones 48 h ACI in WT GFP clones in WT background (blue, from Fig. 2d) quantified with the one cell distance criteria. Same quantification in *FRT40A ubi-nlsGFP/FRT40A bcat-βgal* where *2×βgal* clones were counted as split when clone cells were separated by a single cell and were associated with a continuous group of sibling *2×GFP* cells. This quantification showed no differences with the WT GFP clones in WT background, demonstrating that our method does not produce false positive results. However, it slightly underestimates the total number of fragmented clones (compare with 'all', where every split *2×βgal* clone is counted); n = number of clones. Statistical tests are Fisher's exact tests performed with WT GFP clones in WT background

(blue). **c**, Wing discs 48 h ACI in control (WT in WT) and in supercompetition assay with loser cells expressing β-gal, *UAS-ecad*, *UAS-sqhE20E21*, *UAS-diap1* (an endogenous apoptosis inhibitor) or *UAS-p35* (a bacterial caspase 3 inhibitor) and *fwe^{lose}* RNAi; or after induction of winner clones (*UAS-p35*, *UAS-myc* in WT, *p35* is necessary to block the cell autonomous death induced by high *myc* overexpression[52], and WT in $M^{-/+}$ where WT clones have no GFP). White arrowheads show fragmented clones. Insets show close-up view of representative clones (see Fig. 2d for number of clones analysed). Scale bars, 100 μm. **d**, Scatter plot showing the proportion of fragmented clone (y axis) against the average size of clone (x axis) 48 h ACI for all the different genotypes used in this study (see legend). One dot represents one fragmentation assay. There is no correlation between clone size and clone splitting. Pearson correlation coefficient = 0.14. Note also that without the outlier point (*UAS-pten RNAi* in WT) the correlation is close to 0 (correlation coefficient = −0.036).
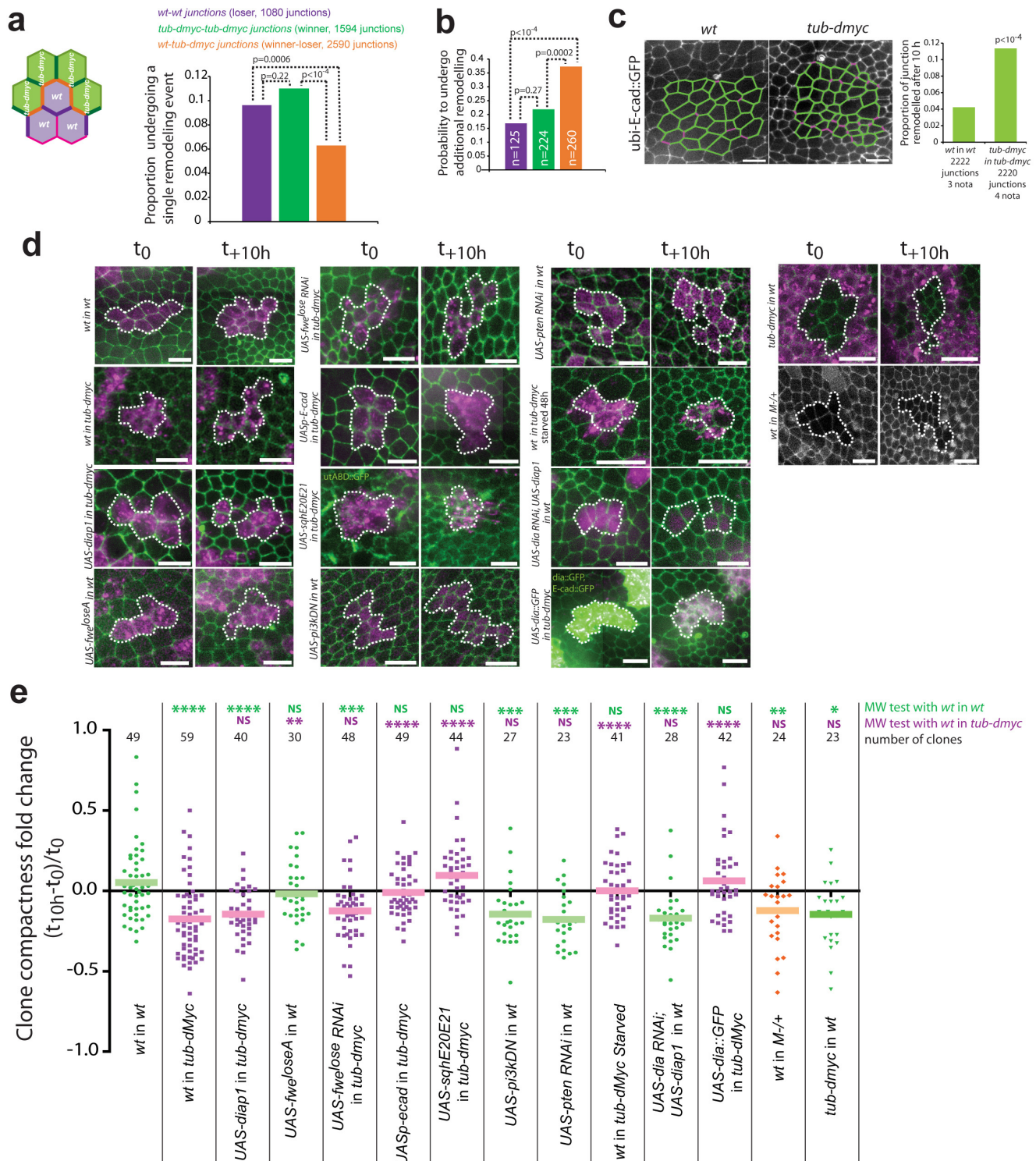
Extended Data Figure 5 | Clone fragmentation is driven by winner–loser mixing. **a**, Left: schematic showing loser cells (WT, purple) and winner cells (*tub-dmyc*, green). Orange junctions are junctions shared by a winner and a loser cell (winner–loser junctions). Dark green junctions are the winner–winner junctions (sharing one vertex with a loser cell) and dark purple are the loser–loser junctions (sharing one vertex with a winner cell) used for the analysis. Right: proportion of junctions undergoing a single remodelling event over 10 h in the notum. *P* values, Fisher's exact tests. **b**, Probability to undergo additional junction remodelling after a first remodelling event. *n* = number of junctions. *P* values, Fisher's exact tests. This suggests that winner–loser junctions undergoing first remodelling events have a higher probability of reverting to the initial topology. **c**, Left: snapshots of WT cells and

tub-dmyc cells in the notum (no clone) at t_0. Purple junctions disappear after 10 h while green junctions remain unchanged (see Supplementary Video 8, representative of three and four videos, respectively). Scale bars, 10 μm. Right: proportion of junction disappearing after 10 h. *P* values, Fisher's exact tests. **d**, Examples of clones in the notum at t_0 and 10 h later for various genotypes. E-cad::GFP is in green and UAS–RFP in purple. The white dotted lines show clone contours. See Extended Data Fig. 5e for the number of clones analysed for each. Scale bars, 10 μm. **e**, Fold change of clone compactness after 10 h in the notum (see Methods). One dot represents one clone. The bars are averages. *P* values, Mann–Whitney tests performed with WT in WT (green) or WT in *tub-dmyc* (purple) (NS, $P > 0.05$; *$P < 0.05$; **$P < 10^{-2}$; ***$P < 10^{-3}$; ****$P < 10^{-4}$).
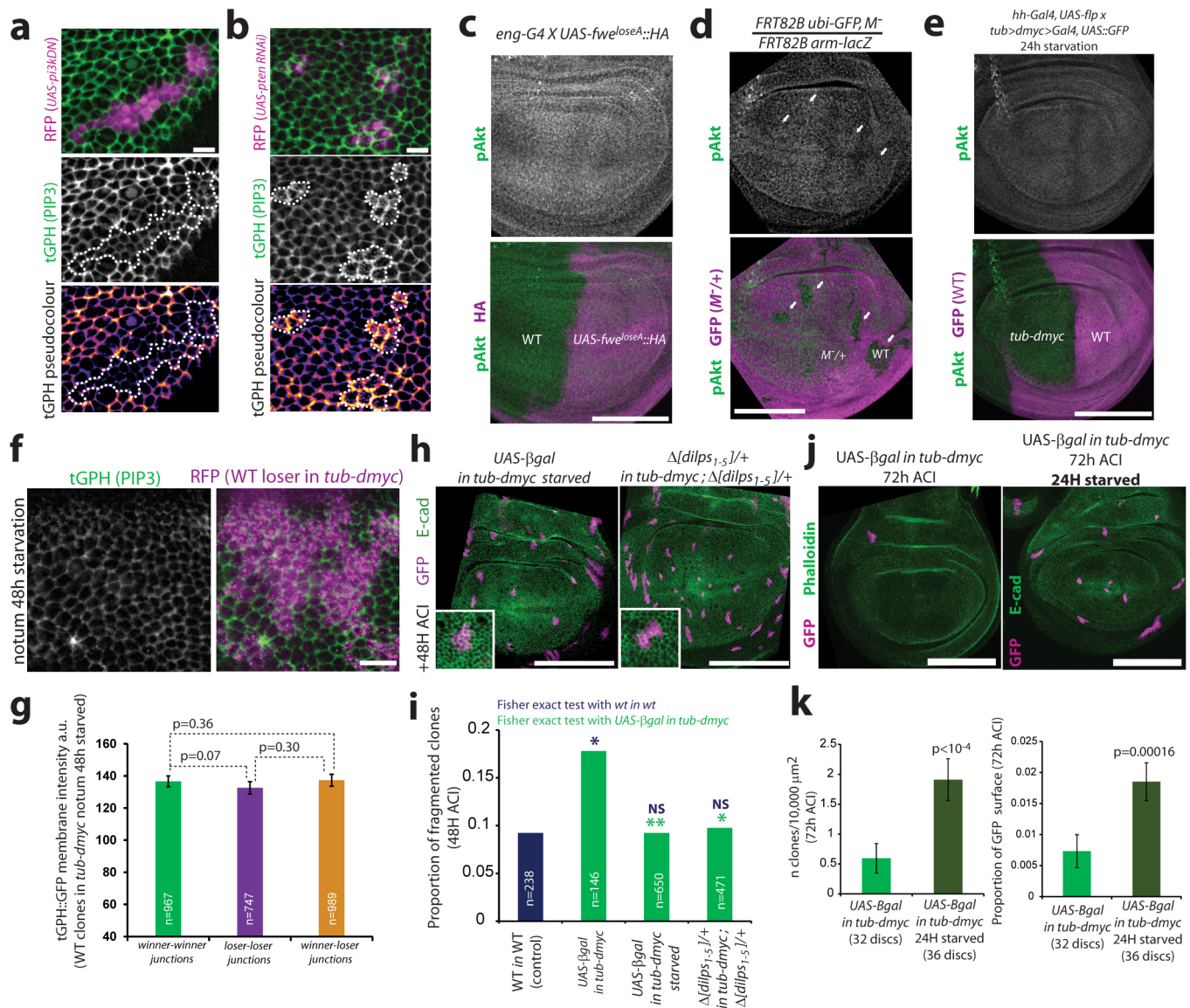
Extended Data Figure 6 | Differential PIP3 drives clone splitting. a, b, The *z*-projection of tGPH (PIP3, green and pseudocolour) in the pupal notum in clones (RFP, purple) overexpressing a dominant negative of PI3K (*UAS-pi3kDN* (**a**), representative of 14 clones) or upon downregulation of PTEN (*UAS-pten* RNAi (**b**), representative of 10 clones). White dotted lines show clone boundaries. Scale bars, 10 μm. **c,** The *z*-projection of a phospho-Akt staining (green) in wing disc overexpressing *fwe*[loseA]::*HA* in the posterior compartment (purple, *eng-G4*, representative of 16 discs). Scale bar, 100 μm. **d,** phosphor-Akt in WT clones (no GFP) surrounded by *M*[−/+] cells (representative of 21 discs). White arrows point to some WT clones. Scale bar, 100 μm. **e,** The *z*-projections of phospho-Akt (green), GFP (magenta) in wing discs after removal of one additional copy of *myc* in the posterior compartment and 24 h of starvation (*hh-gal4, UAS-flp* × *tub>dmyc>gal4*, *UAS–GFP*, representative of ten discs). Scale bar, 100 μm. **f,** The *z*-projection of tGPH (PIP3) in loser clones (supercompetition assay, purple are losers) in the pupal notum after 48 h of starvation. Representative of 28 clones. Scale bars, 10 μm. **g,** Quantification of the mean junction membrane intensity of tGPH in winner–winner, loser–loser and winner–loser junctions in the notum; *n* = number of junctions. Error bars, s.e.m. *P* values, Mann–Whitney tests. **h,** Wing discs with loser cells (supercompetition assay) 48 h ACI after 24 h of starvation, or upon removal of one copy of *Drosophila* insulin-like peptides (*Dilp1* to *Dilp5*). Insets show representative clones of 650 and 471 clones, respectively. Scale bars, 100 μm. **i,** Proportion of fragmented clones; *n* = number of clones. Statistical tests are Fisher's exact tests performed with WT (in blue, WT in WT) or control competition (in green, *UAS-βgal* in *tub-dmyc*). NS, *P* > 0.05; **P* < 0.05; **P < 10[−2]. WT in WT and WT in *tub-dmyc* come from Fig. 2d. This result suggests that losers and winners have differential abilities in processing and responding to extracellular insulin. **j,** Wing discs showing *UAS-βgal* clones (loser, purple) in *tub-dmyc* background 72 h ACI with or without 24 h of starvation, representative of 32 and 36 discs. Scale bars, 100 μm. **k,** Average density of loser clones (left) and average proportion of the wing pouch covered with GFP-positive cells (right) 72 h ACI. Error bars, s.e.m. *P* values, Mann–Whitney tests.
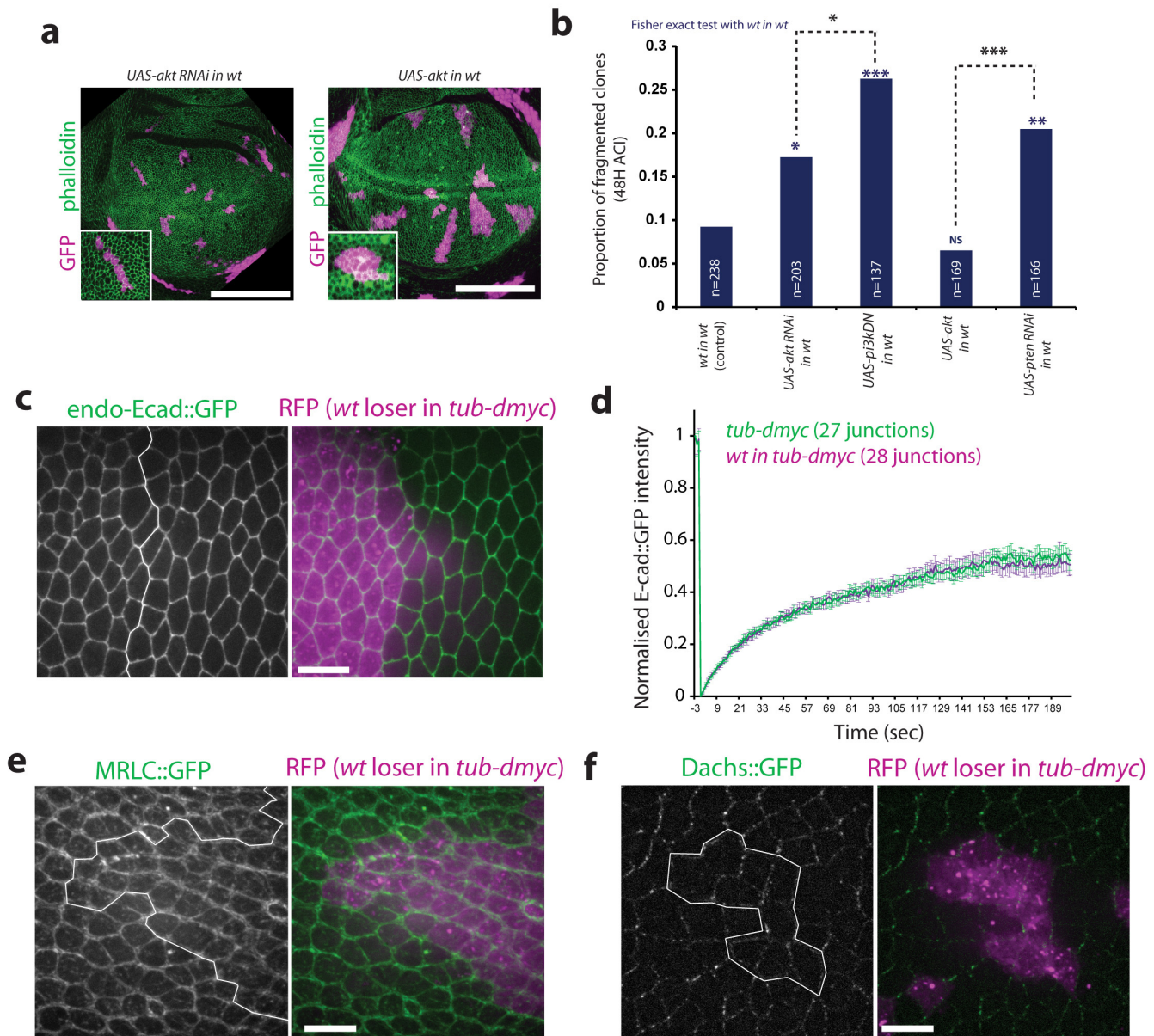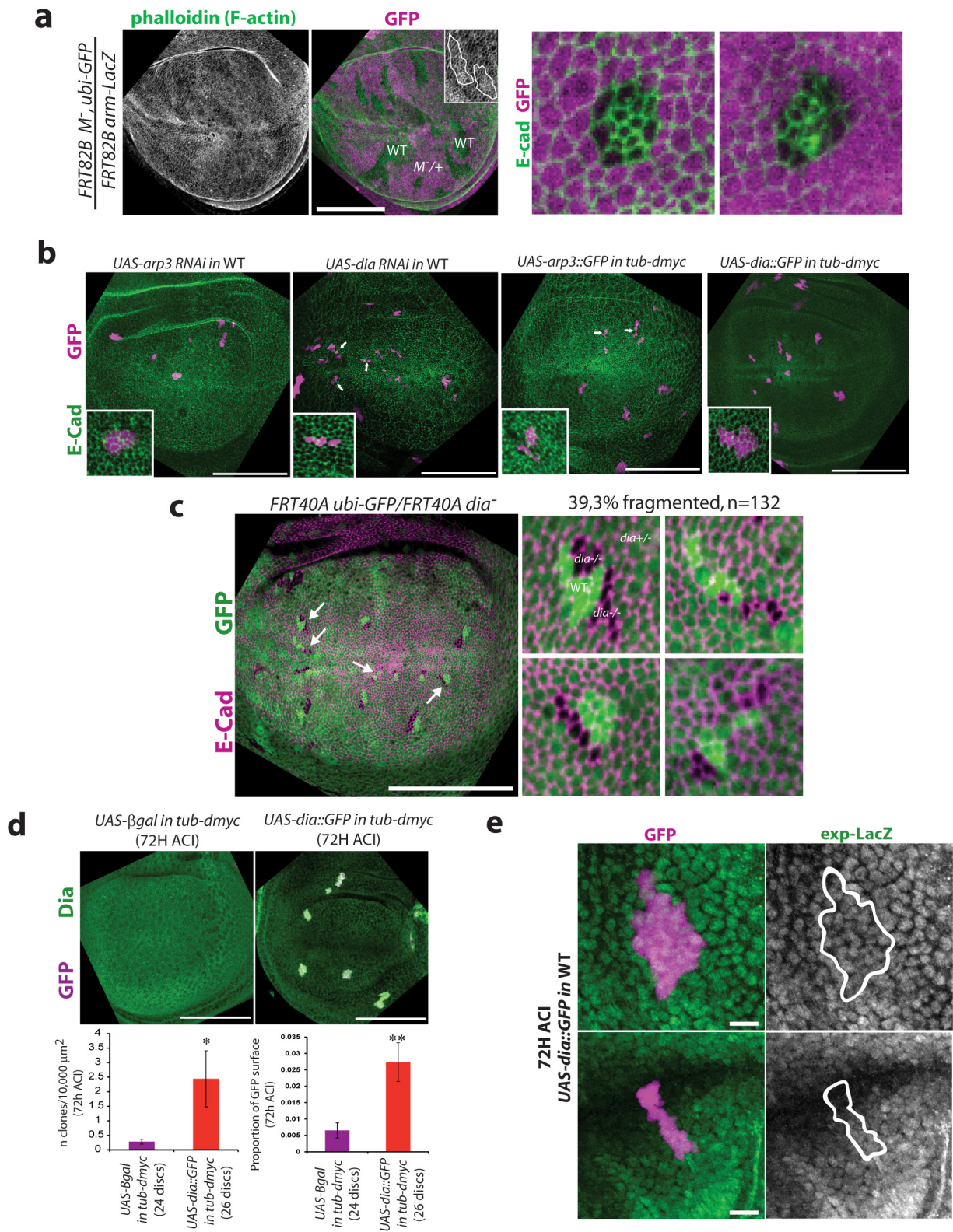
Extended Data Figure 7 | Akt is not sufficient to explain winner–loser mixing, and E-cad, myosin II and Dachs do not show visible defects in loser clones. a, Wing discs showing clones upon downregulation of Akt (*UAS-akt RNAi*, left) or upregulation of Akt (*UAS-akt*, right) 48 h ACI. Insets show close-up views of a representative clone in each condition for 203 and 169 clones, respectively. Scale bars, 100 μm. **b**, Proportion of fragmented clones. WT in WT comes from Fig. 2d. *UAS-pi3kDN* in WT and *UAS-pten RNAi* in WT come from Fig. 3d. e. *n* = number of clones. Statistical tests are Fisher's exact tests performed with WT (in blue, WT in WT) or as indicated by the dotted lines. NS, $P > 0.05$; *$P < 0.05$; **$P < 10^{-2}$; ***$P < 10^{-3}$. **c**, The *z*-projection of endo-Ecad::GFP (knock-in line) in the pupal notum. Loser clones are marked with RFP (WT in *tub-dmyc* background, representative of 30 clones and 3 nota). White line marks the clone contour. Scale bar, 10 μm. **d**, Average

normalized intensity recovery curves of endo-Ecad::GFP after photobleaching in loser–loser junctions (WT in *tub-dmyc*, purple) and in winner–winner junctions (*tub-dmyc*) in the notum. Error bars, s.e.m. **e**, The *z*-projection of MRLC::GFP (endogenous promoter, *spaghetti-squashed*, *sqh*) in the pupal notum. Loser clones are marked with RFP (WT in *tub-dmyc* background, representative of 30 clones and 5 nota). White line marks the clone contour. Scale bar, 10 μm. Note that utABD::GFP (as shown in Fig. 4a) is under the control of the same promoter, thus the actin reduction in losers is not due to a reduction of *sqh* promoter activity. **f**, The *z*-projection of Dachs::GFP (endogenous promoter) in the pupal notum. Loser clones are marked with RFP (WT in *tub-dmyc* background, representative of 17 clones and 2 nota). White line marks the clone contour. Scale bar, 10 μm.
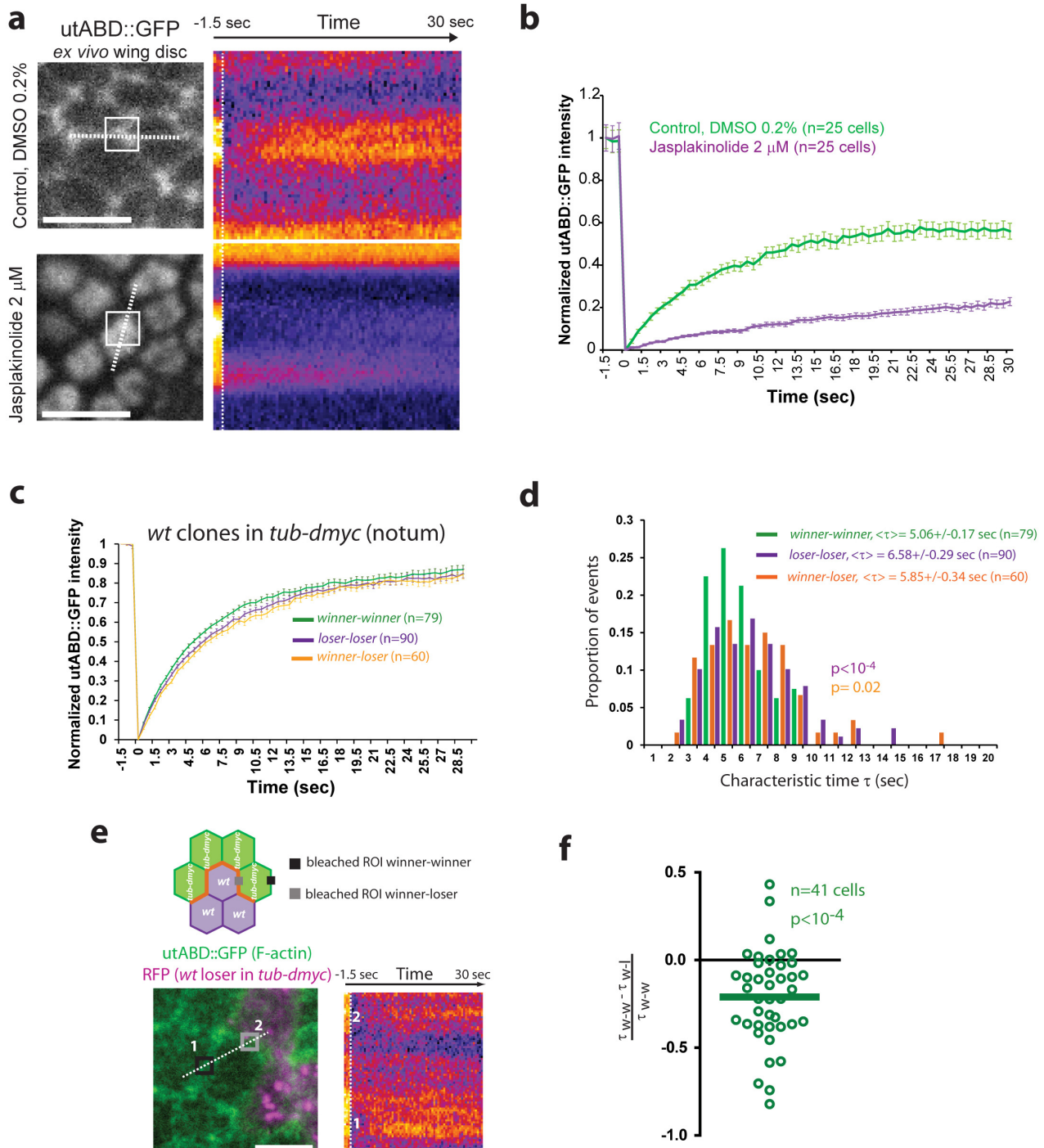
Extended Data Figure 8 | F-actin turnover/polymerization rate is reduced in loser junctions. The utABD::GFP has been previously used to assess actin dynamics[39,49]. This is further demonstrated by the experiment described in **a** and **b**. **a**, *Ex vivo* cultured wing discs in control media (DMSO 0.2%) or in media containing 2 μM jasplakilonide, an inhibitor or actin turnover[53]. White rectangles are the bleached zones. White dotted line is that used for the kymographs shown on the right; $t = 0$ s (white dotted line on kymographs) is the time of bleaching. Images representative of 25 FRAP experiments for each condition. Scale bars, 5 μm. **b**, Averaged normalized recovery curves of utABD::GFP intensity after photobleaching in control and jasplakinolide-treated wing discs. Error bars, s.e.m. **c**, Averaged utABD::GFP normalized intensity recovery curves in loser–loser junctions (purple), winner–winner junctions (green) and winner–loser junctions (orange) after photobleaching (WT losers in *tub-dmyc*) in the notum. Error bars, s.e.m. **d**, Distribution of the characteristic times of utABD::GFP intensity recovery in winner–winner, loser–loser and winner–loser junctions. *P* values, Mann–Whitney tests. **e**, Top, schematic of the FRAP experiments; two ROIs are bleached simultaneously in the same winner cell (*tub-dmyc*) sharing contacts with a loser cell (WT). Grey square, winner–loser bleached junction; black square, winner–winner bleached junction. Bottom, confocal image in the pupal notum 48 h ACI of utABD::GFP in a supercompetition assay (purple cells, losers). Scale bar, 5 μm. Squares show the simultaneously bleached regions (1, winner–winner junction; 2, winner–loser junction); the white dotted line is the line used for the kymograph shown on the right, representative of 41 FRAP experiments. **f**, Distribution of the fold change of the characteristic time of intensity recovery in the winner–winner junction compared with a winner–loser junction of the same cell ($(\tau_{w-w} - \tau_{w-l})/\tau_{w-w}$). One dot represents one cell. The bar is the average. The statistical test is a one-sample *t*-test with 0 as reference value.

a FRT82B M⁻, ubi-GFP / FRT82B arm-LacZ

phalloidin (F-actin)　　GFP　　　　E-cad GFP

WT　WT
M⁻/+

b UAS-arp3 RNAi in WT　UAS-dia RNAi in WT　UAS-arp3::GFP in tub-dmyc　UAS-dia::GFP in tub-dmyc

GFP

E-Cad

c FRT40A ubi-GFP/FRT40A dia⁻　　39,3% fragmented, n=132

GFP

E-Cad

dia+/-
dia-/-
WT
dia-/-

d UAS-βgal in tub-dmyc (72H ACI)　UAS-dia::GFP in tub-dmyc (72H ACI)

Dia

GFP

n clones/10,000 µm² (72h ACI)

*

UAS-Bgal in tub-dmyc (24 discs)
UAS-dia::GFP in tub-dmyc (26 discs)

Proportion of GFP surface (72h ACI)

**

UAS-Bgal in tub-dmyc (24 discs)
UAS-dia::GFP in tub-dmyc (26 discs)

e 72H ACI UAS-dia::GFP in WT

GFP　　　　exp-LacZ

Extended Data Figure 9 | Filamentous actin defects are necessary and sufficient to drive clone fragmentation. a, The *z*-projection of phalloidin (green) and GFP (magenta) in a wing disc containing WT clones (no GFP) in $M^{-/+}$ background representative of 20 discs. Top inset shows phalloidin signal for two WT clones (white lines). Right: close-up views of cell shape in two WT clones. Scale bar, 100 μm. **b**, Wing discs 48 h ACI upon silencing of Arp3 (*arp3 RNAi*) in WT background, in supercompetition assay with loser cells expressing Arp3::GFP, silencing of Dia (*dia RNAi*) in WT background, and in supercompetition assay with loser cells expressing Dia::GFP. Discs correspond to experiments quantified in Fig. 4d; control WT and control supercompetition assays are the same as in Fig. 2d. White arrowheads show fragmented clones. Insets show close-up view of a representative clone in each condition out of 108, 174, 259 and 233 clones analysed, respectively. Scale bars, 100 μm. **c**, Wing disc 48 h ACI. The clones without GFP are homozygous mutant for *diaphanous* (*dia⁵*, hypomorphic allele[40]), the sibling WT clones have two copies of GFP. Four close-up views of fragmented mutant clones are shown on the right, representative of 132 clones analysed. The proportion of fragmented clones is 39.3%, counting every fragmented clone (including patches separated by more than one cell). Scale bar, 100 μm. **d**, Top: representative wing discs during supercompetition 72 ACI in loser cells (purple, GFP) overexpressing β-gal (control, representative of 24 discs) or Dia::GFP (representative of 26 discs). Scale bar, 100 μm. Bottom: quantification of the mean loser clone density and the average proportion of the wing pouch surface covered by loser clones. *$P < 0.05$; **$P < 10^{-2}$; Mann–Whitney test. Error bars, s.e.m. **e**, Level of expression of *expanded* (*exp-lacZ*), a downstream target of Yki/YAP, in two representative examples of clones overexpressing Dia::GFP (*hs-flp22, act<y<gal4*; *UAS–GFP × UAS-dia::GFP*) 72 h ACI (representative of 31 clones out of 16 wing discs). White lines show the contour of the clones. Scale bars, 5 μm.
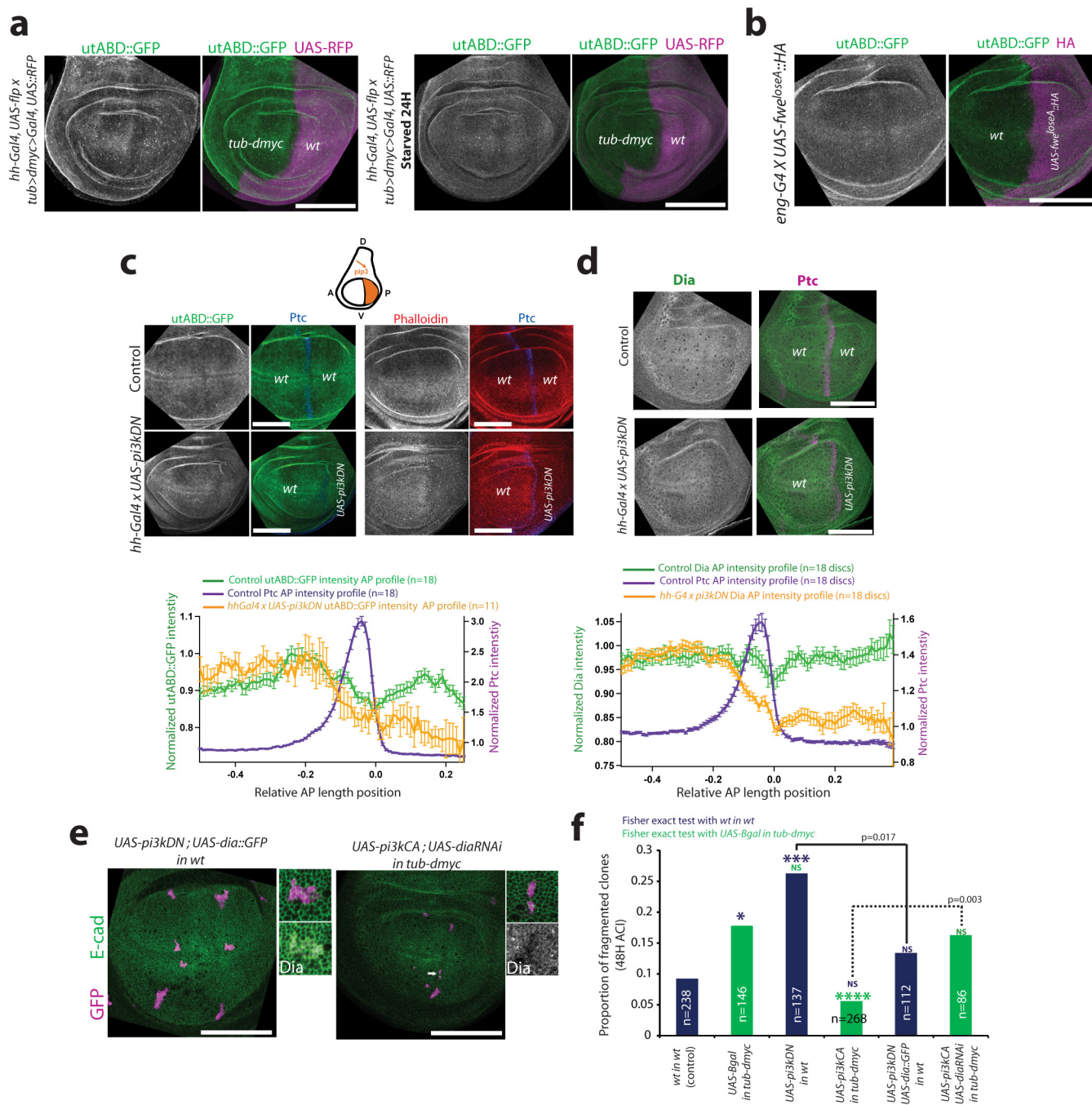
Extended Data Figure 10 | PIP3 acts upstream of actin defects. Starvation was sufficient to abolish differences in F-actin between compartments expressing different levels of *myc* (Extended Data Fig. 10a), whereas overexpression of *fwe*^{*loseA*} in the posterior compartment did not modify F-actin (Extended Data Fig. 10b). Moreover, PIP3 downregulation in a full compartment was sufficient to downregulate actin (Extended Data Fig. 10c) and junctional Dia (Extended Data Fig. 10d). Finally, overexpression of Dia significantly reduced the number of fragmented clones upon downregulation of PIP3, whereas knocking down Dia and increasing PIP3 impaired the rescue of loser fragmentation (Extended Data Figure 10e, f, $P = 0.017$ and 0.003, respectively). Altogether, we concluded that actin defects are driven by the modulation of PIP3 in loser cells. **a**, The *z*-projections of utABD::GFP (green, F-actin) and RFP (magenta) in wing discs after removal of one additional copy of *myc* in the posterior compartment (*hh-gal4, UAS-flp × tub>dmyc>gal4, UAS–RFP*, left, representative of 20 discs) and after 24 h of starvation (right, representative of 9 discs out of 10). Scale bars, 100 μm. **b**, The *z*-projection of utABD::GFP (green, F-actin) in a wing disc overexpressing *fwe*^{*loseA*}::HA in the posterior compartment (purple, *eng-G4*, representative of 20 discs). Scale bar, 100 μm. **c**, Stainings of utABD::GFP and phalloidin in control wing discs (*hh-gal4* alone, representative of 18 discs) and upon reduction of PIP3 in the posterior compartment using *hh-gal4* and *UAS-pi3kDN* (see top scheme: A, anterior; P, posterior; D, dorsal; V, ventral, representative of 11 discs). The posterior compartment is at the right side of the Patched (Ptc) stripe marked in blue. Scale bars, 100 μm. Bottom: averaged normalized intensity line profiles along the antero-posterior axis for utABD::GFP (green and yellow) and Ptc (blue). Position 0 corresponds to the maximum inflexion of Ptc intensity peak (right side of the stripe); *n* = number of wing discs. Error bars, s.e.m. **d**, Dia staining in control (*hh-gal4* alone, representative of 18 discs) and upon reduction of PIP3 in the posterior compartment using *hh-gal4* and *UAS-pi3kDN* (representative of 18 discs). Scale bars, 100 μm. Bottom: averaged normalized intensity line profile taken along the antero-posterior axis for Dia (green and yellow) and Ptc (blue); *n* = number of wing discs. Error bars, s.e.m. **e**, Wing discs 48 h ACI showing clones (*hs-flp22; act<y<gal4; UAS–GFP* purple) in WT background overexpressing a dominant negative form of PI3K (*UAS-pi3kDN*) and *dia::GFP*, or loser clones in *tub-dmyc* background overexpressing a constitutively active form of PI3K (*pi3kCA*) and *dia RNAi*. The white arrowhead shows a fragmented clone. Insets show close-up views of a representative clone out of 112 and 86 clones, respectively. Scale bars, 100 μm. **f**, Proportion of fragmented clones. Blue bars are *act<y<Gal4* clones in WT background; green bars are loser clones in *tub-dmyc* background. WT in WT and *UAS-βgal* in *tub-dmyc* (supercompetition assay) come from Fig. 2d. *UAS-pi3kDN* in WT and *pi3kCA* in *tub-dmyc* come from Fig. 3d, e; *n* = number of clones. Statistical tests are Fisher's exact tests performed with WT in WT (in blue) or *UAS-βgal* in *tub-dmyc* (in green) or as indicated by the black dotted lines. NS, $P > 0.05$; *$P < 0.05$; ***$P < 10^{-3}$; ****$P < 10^{-4}$.

LETTER

A cytosolic network suppressing mitochondria–mediated proteostatic stress and cell death

Xiaowen Wang[1] & Xin Jie Chen[1]

Mitochondria are multifunctional organelles whose dysfunction leads to neuromuscular degeneration and ageing. The multifunctionality poses a great challenge for understanding the mechanisms by which mitochondrial dysfunction causes specific pathologies. Among the leading mitochondrial mediators of cell death are energy depletion, free radical production, defects in iron–sulfur cluster biosynthesis, the release of pro-apoptotic and non-cell-autonomous signalling molecules, and altered stress signalling[1–5]. Here we identify a new pathway of mitochondria-mediated cell death in yeast. This pathway was named mitochondrial precursor over-accumulation stress (mPOS), and is characterized by aberrant accumulation of mitochondrial precursors in the cytosol. mPOS can be triggered by clinically relevant mitochondrial damage that is not limited to the core machineries of protein import. We also discover a large network of genes that suppress mPOS, by modulating ribosomal biogenesis, messenger RNA decapping, transcript-specific translation, protein chaperoning and turnover. In response to mPOS, several ribosome-associated proteins were upregulated, including Gis2 and Nog2, which promote cap-independent translation and inhibit the nuclear export of the 60S ribosomal subunit, respectively[6,7]. Gis2 and Nog2 upregulation promotes cell survival, which may be part of a feedback loop that attenuates mPOS. Our data indicate that mitochondrial dysfunction contributes directly to cytosolic proteostatic stress, and provide an explanation for the association between these two hallmarks of degenerative diseases and ageing. The results are relevant to understanding diseases (for example, spinocerebellar ataxia, amyotrophic lateral sclerosis and myotonic dystrophy) that involve mutations within the anti-degenerative network.

ANT1 (also known as *SLC25A4*) encodes adenine nucleotide translocase involved in ATP/ADP exchange across the mitochondrial inner membrane. Mutations in the human *ANT1* gene cause autosomal dominant progressive external ophthalmoplegia, cardiomyopathy and myopathy[8,9]. In yeast, similar mutations in the orthologous *AAC2* (also known as *PET9*) gene cause protein misfolding, followed by severe mitochondrial damage and ageing-dependent degenerative cell death[10,11]. We used the dominant $AAC2^{A128P}$ allele to determine how mitochondrial damage causes cell death. We screened for multicopy suppressors that inhibit $AAC2^{A128P}$-induced cell death (Fig. 1a). From 336 viable transformants, 40 genes capable of suppression were identified (Fig. 1b and Supplementary Table 1). Thirty-two of them fall into one of the five functional categories: target of rapamycin (TOR) signalling, mRNA silencing/turnover, ribosomal function/protein translation, transfer RNA methylation, and cytosolic protein chaperoning/degradation (Fig. 1c). These anti-degenerative cell death genes belong to a functional network that controls cytosolic protein homeostasis. The remaining eight include the uncharacterized *YEL057C*, *YMR074C*, *YOL098C* and the truncated *YPR022C* that are named *SDD1–4* (suppressor of degenerative death), respectively (Fig. 1d).

Among the suppressors are Tod6, Dot6 and Rpd3 that repress ribosomal biogenesis in the TOR signalling pathway[12]. Dcp2 and Edc3 decap mRNAs, silence cap-dependent translation and promote mRNA degradation. Tif4632, Tif3, Cdc33 and Prt1 are translation initiation factors, with Tif4632 also being known to activate cap-independent translation[13]. Pbp1 and Psp2 are constituents of P-bodies and stress granules that process and silence mRNAs under stress conditions[14]. Nam7 participates in nonsense-mediated mRNA decay. Poc4 and Ump1 promote the assembly and maturation of the proteasome. Ssb1, Ssb2 and Zuo1 are ribosome-associated chaperones.

Notably, the suppressors include seven ribosomal proteins and two tRNA methyltransferases, suggesting a link between mitochondrial stress and cytosolic translation. Rpp0 mediates interactions between translational elongation factors and the ribosome and its overexpression cures prions[15]. The Rpl40A and Rpl40B paralogues and Rpl38 promote transcript-specific translation in yeast and humans[16,17]. The Trm9 methyltransferase specifically methylates the wobble uridine residues (underlined) in the anticodons of $tRNA^{GLU(\underline{U}UC)}$ and $tRNA^{ARG(\underline{U}CU)}$, which stimulates the translation of stress-responsive proteins encoded by mRNAs enhanced in these two specific codons[18]. In total, 11 out of the 40 suppressors are Trm9-skewed proteins (Fig. 1c, asterisks), suggesting that suppression by Trm9 may involve increased translation of these proteins.

The suppressor genes also promote cell viability in mutants with diverse mitochondrial damage. Deletion of *YME1* encoding the i-AAA chaperone/protease on the inner membrane[19], and *ATP1* encoding the α-subunit of F_1-ATPase, are lethal when mitochondrial DNA (mtDNA) is eliminated (termed ρ^o) by ethidium bromide. This lethality is mitigated by overexpressing the anti-degenerative genes (Extended Data Fig. 1a).

All of the suppressors are located in the cytosol, suggesting that they prevent cell death via pathways that operate in the cytosol. $AAC2^{A128P}$ expression and mtDNA elimination from the *yme1* and *atp1* mutants collapse the mitochondrial membrane potential $\Delta\psi_m$ that is critical for protein import[20]. We thus propose that mitochondrial damage in these strains affects protein import, leading to the over-accumulation and misfolding of mitochondrial precursors in the cytosol. The resulting stress, which we term mitochondrial precursor over-accumulation stress (mPOS), would then trigger cell death (Extended Data Fig. 1b). The anti-degenerative suppressor genes might reverse mPOS by reducing translation, thus decreasing stress imposed by protein over-accumulation, misfolding and aggregation. Or, they might selectively stimulate translation of stress-resistant proteins, and increase specific chaperoning activity and protein turnover.

Several lines of evidence support the mPOS hypothesis. First, the anti-degenerative genes suppressed ρ^o-lethality in the *tom70Δ* and *mgr2Δ* mutants that reduce protein translocation across the outer and inner mitochondrial membranes, respectively (Extended Data Fig. 1a). Overexpression of the mPOS suppressors *PBP1* and *SSB1* is known to rescue ρ^o-lethality of *tim18* mutants defective in the TIM22 protein translocase[21]. Second, we detected synthetic growth defects between *yme1Δ* and disruption mutations in anti-degenerative genes including *PBP1*, *RPL40A*, *SSB1* and *POC4* (Extended Data Fig. 1c). This is consistent with a crosstalk between mitochondrial and cytosolic proteostatic functions. Synthetic growth defects were also observed between

[1]Department of Biochemistry and Molecular Biology, State University of New York Upstate Medical University, Syracuse, New York 13210, USA.
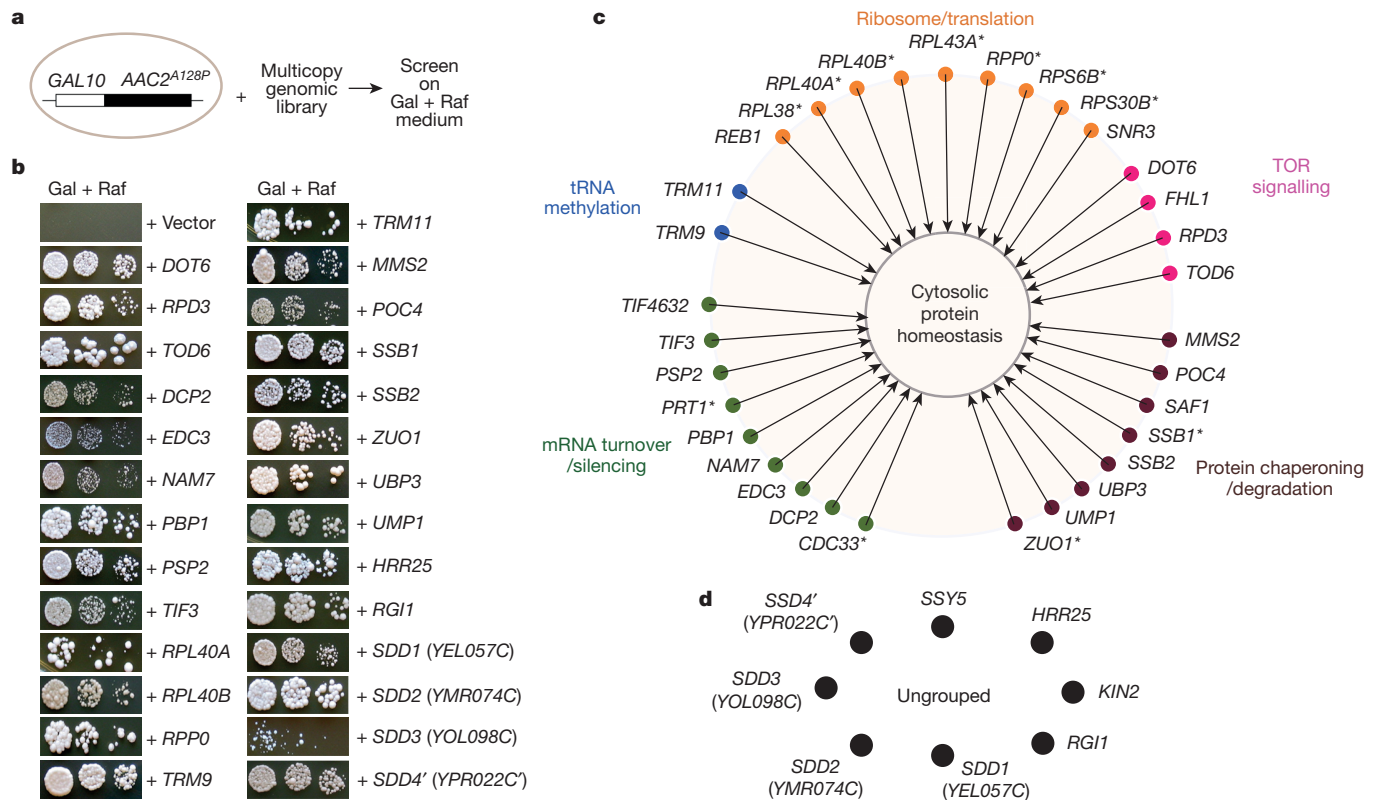
Figure 1 | A cytosolic anti-degenerative network that suppresses mitochondria-induced cell death. a, Schematic of multicopy suppressor screen. **b**, Suppression of $AAC2^{A128P}$-induced cell death on galactose (Gal) plus raffinose (Raf) medium. **c**, The anti-degenerative network that improves cytosolic protein homeostasis and cell survival in $AAC2^{A128P}$-expressing cells. The asterisks and prime symbols in **b** and **c** denote Trm9-skewed and -truncated proteins, respectively. **d**, Ungrouped anti-degenerative genes.

$yme1\Delta$ and the disruption of genes involved in cytosolic mRNA decay ($DHH1$) and proteasomal biogenesis ($BLM10$ and $RPN4$) (Extended Data Fig. 1c). Third, expression of the aggregation-prone human Huntingtin (HD) protein with a polyglutamine track (25Q) was tolerated in wild-type but not in $AAC2^{A128P}$ cells (Extended Data Fig. 2a). Aggregation of HD(25Q) was frequently detected in ρ^o but not in ρ^+ cells (Extended Data Fig. 2b). Fourth, $AAC2^{A128P}$ expression increased the accumulation and aggregation of the mitochondrial Aco1 protein in the cytosol (Extended Data Fig. 2c, d). Thus, mitochondrial damage aggravates protein aggregation in the cytosol.

Finally, we used mass spectrometry to compare the cytosolic proteomes from the wild-type and $AAC2^{A128P}$ cells. The $AAC2^{A128P}$ mutant survives poorly at 25 °C and has a significantly reduced protein synthesis rate (Extended Data Fig. 3). A total of 1,923 proteins were detected in each strain (Supplementary Note). $AAC2^{A128P}$ expression downregulated 107 proteins by >2-fold, with the majority being metabolic enzymes (Supplementary Table 2). There were 297 proteins upregulated by >1.266-fold ($P < 0.05$) in $AAC2^{A128P}$ cells (Extended Data Fig. 4).

Among the 43 proteins with >2-fold upregulation, 24 (55.8%) are nuclear-encoded mitochondrial proteins (Fig. 2 and Supplementary Table 3). The cytosolic accumulation of several proteins was confirmed by western blot (Extended Data Fig. 5a). These proteins are unimported precursors rather than proteolytically matured proteins released from damaged mitochondria, because their presequences are detected by mass spectrometry (Extended Data Fig. 6). Together, these data reveal cytosolic accumulation of mitochondrial precursors, which we propose triggers mPOS.

Mitochondrial damage is known to abrogate the heat-shock response[22]. Consistent with this, we found that mPOS did not upregulate the major heat-shock-inducible Hsp70-type chaperones (for example, Ssa1, Ssa2, Sse1 and Sse2; Extended Data Fig. 7). Upregulation of these chaperones would probably exacerbate mPOS by promoting the fold-ing of mitochondrial precursors, which would interfere with import and/or block cytosolic degradation. We also found that the levels of Ssb1 and Ssb2, the holdase-type Hsp70-family chaperones that cotranslationally bind to and stabilize structurally disordered and aggregation-prone nascent polypeptides[23,24], were not induced in the $AAC2^{A128P}$ mutant (Extended Data Fig. 7 and Supplementary Note). However, 1–2 extra copies of $SSB1$ are sufficient to suppress the heat sensitivity of the $yme1\Delta$ mutant (Extended Data Fig. 8). These specific chaperones may maintain the unimported precursors in an unfolded state thereby preventing aggregation. In support of this, we found that the mitochondrial proteins Aco1 and Abf2 are substrates of Ssb1 in the cytosol and that their association with Ssb1 is increased in the $AAC2^{A128P}$-expressing cells (Extended Data Fig. 5b).

Nineteen non-mitochondrial proteins were upregulated by >2-fold in $AAC2^{A128P}$-expressing cells (Fig. 2). We speculated that the three highly conserved ribosome-associated proteins, Gis2, Nog2 and Tma7, may be upregulated to benefit cell viability. Indeed, overexpression of these genes is sufficient to rescue $AAC2^{A128P}$-induced cell death (Fig. 3a). $GIS2$ overexpression also rescued ρ^o-lethality of the $atp1\Delta$, $mgr2\Delta$ and $tom70\Delta$ mutants (Extended Data Fig. 9a). Overexpression of Gis2, Nog2 and Tma7 did not reduce global protein synthesis (Extended Data Fig. 9b). Instead, these proteins probably promote cell survival by stimulating the translation of specific mRNAs. Indeed, Gis2 and its human orthologue ZNF9 (also known as CNBP) are proposed to stimulate cap-independent translation[6].

Nog2 inhibits nuclear export of the 60S ribosomal subunit[7]. Consistent with this, we found that Nog2 overexpression increased the total levels of free 60S but not the 40S ribosomal subunit, concomitant with a decrease of polysomes (Fig. 3b). Possibly, nuclear retention of 60S and delayed maturation into the 80S in the cytosol remodel translational programs in favour of cell viability. Decreased availability of 60S is known to benefit the survival of ageing cells[10,25]. The free 60S
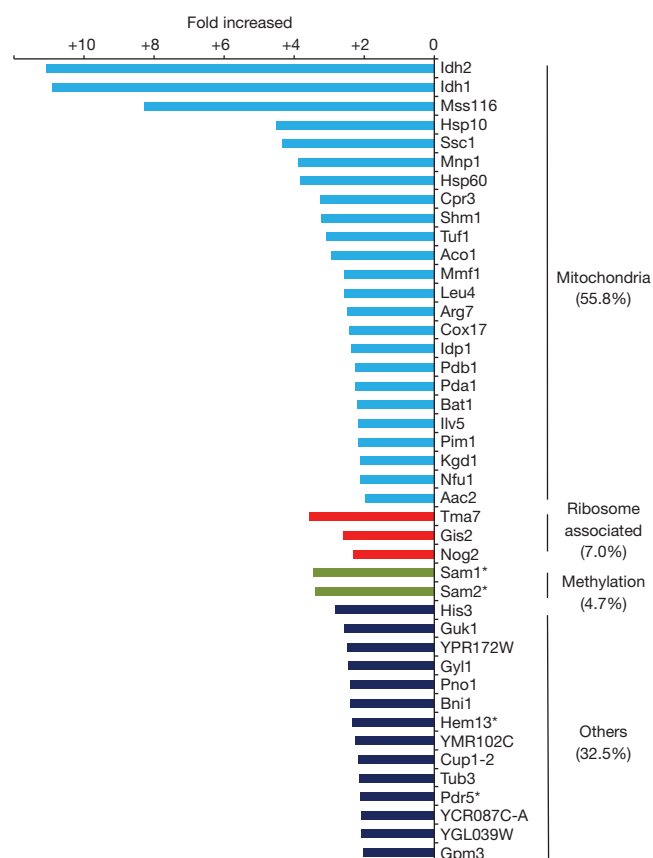
Figure 2 | Comparison of cytosolic proteomes from *AAC2^{A128P}* and wild-type cells. Isobaric tag for relative and absolute quantitation mass spectrometry (iTRAQ) analysis showing 43 proteins that are upregulated by >2-fold in the cytosol of *AAC2^{A128P}* cells. Asterisk denotes Trm9-skewed non-mitochondrial proteins.

was also increased by *AAC2^{A128P}* expression, which may be at least partially attributed to *NOG2* overexpression.

Gis2 and Nog2 can also be induced by the mitochondrial uncoupler carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) (Fig. 3c), and by
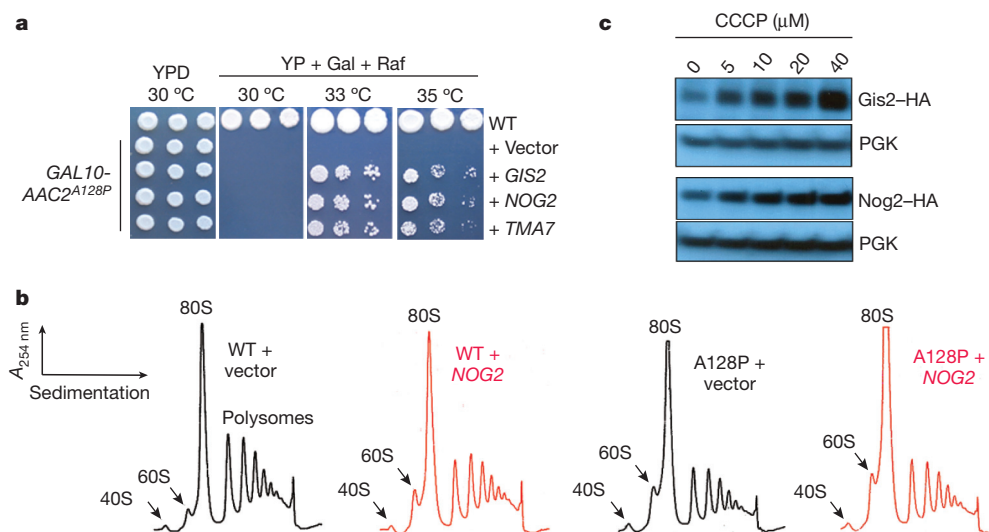

Figure 4 | Gis2 and Nog2 upregulation in response to mitochondrial damage provides a feedback loop to suppress mPOS and promote cell survival. Activation of other pathways in the anti-degenerative network as shown in Fig. 1c, d may also benefit cell survival.

the disruption of *POC4*, *UMP1* and *RPN4*, which affects proteasome biogenesis and assembly (Extended Data Fig. 10a, b). Gis2 and Nog2 are intrinsically unstable and are stabilized in *AAC2^{A128P}* cells (Extended Data Fig. 10c, d). Gis2 is not ubiquitinated (Extended Data Fig. 10e, f), suggesting that it is degraded in a proteasome-dependent but ubiquitin-independent manner. These data are consistent with the idea that Gis2 and Nog2 stabilization helps to keep the mPOS pathway in check.

In summary, we have identified a novel pathway of mitochondria-induced cell death. This pathway, mPOS, is triggered not only by mutations directly affecting the core protein import machineries, but also by conditions that interfere with mitochondrial inner membrane integrity and function. The mPOS model is consistent with recent findings that mitochondrial dysfunction stabilizes cytosolic proteins[26,27] and reduces the mitochondrial import of signalling proteins[28]. Moreover, decreased TOR signalling, which attenuates the progression of mitochondrial disease[29], may do so via a reduction in protein synthesis. The mPOS pathway could be particularly relevant in cells with high mitochondrial density. In these cells, a subfraction of mitochondria with a low $\Delta\psi_m$ may be sufficient to disturb cytosolic protein homeostasis independent of the cell's bioenergetic state. Finally, we identified a network of anti-degenerative genes that protect cells against mPOS, including Gis2 and Nog2 that are intrinsically upregulated in response to mitochondrial stress (Fig. 4). The human orthologues of Gis2 and Nog2, and several other anti-degenerative genes that we identified are involved in neuro-muscular degenerative disorders (Supplementary Table 4). In light of this study, we predict that conditions that promote mPOS may contribute to the pathogenesis of these diseases.

Figure 3 | Upregulation of Gis2 and Nog2 in response to mitochondrial damage promotes cell survival. **a**, Growth inhibition by *AAC2^{A128P}* expressed from the *GAL10* promoter is suppressed by *GIS2*, *NOG2* and *TMA7* on the pRS425 multicopy vector at 33 °C and 35 °C, but not at 30 °C. YP, yeast complete medium; YPD, yeast complete medium with 2% glucose.
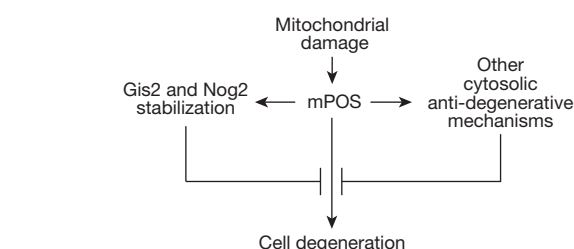
b, Ribosomal profiles of wild-type (WT) and *AAC2^{A128P}* (A128P) cells with or without the overexpression of *NOG2*. **c**, Over-accumulation of haemagglutinin (HA)-tagged Gis2 and Nog2 in response to CCCP treatment. PGK, phosphoglycerate kinase. Full scans of western blots are available in the Supplementary Information.

1. Wallace, D. C. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu. Rev. Genet.* **39**, 359–407 (2005).
2. Veatch, J. R., McMurray, M. A., Nelson, Z. W. & Gottschling, D. E. Mitochondrial dysfunction leads to nuclear genome instability via an iron-sulfur cluster defect. *Cell* **137**, 1247–1258 (2009).
3. Wang, X. The expanding role of mitochondria in apoptosis. *Genes Dev.* **15**, 2922–2933 (2001).
4. Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144**, 79–91 (2011).
5. Pan, Y., Schroeder, E. A., Ocampo, A., Barrientos, A. & Shadel, G. S. Regulation of yeast chronological life span by TORC1 via adaptive mitochondrial ROS signaling. *Cell Metab.* **13**, 668–678 (2011).
6. Gerbasi, V. R. & Link, A. J. The myotonic dystrophy type 2 protein ZNF9 is part of an ITAF complex that promotes cap-independent translation. *Mol. Cell. Proteomics* **6**, 1049–1058 (2007).
7. Matsuo, Y. *et al.* Coupled GTPase and remodelling ATPase activities form a checkpoint for ribosome export. *Nature* **505**, 112–116 (2014).
8. Kaukonen, J. *et al.* Role of adenine nucleotide translocator 1 in mtDNA maintenance. *Science* **289**, 782–785 (2000).
9. Palmieri, L. *et al.* Complete loss-of-function of the heart/muscle-specific adenine nucleotide translocator is associated with mitochondrial myopathy and cardiomyopathy. *Hum. Mol. Genet.* **14**, 3079–3088 (2005).
10. Wang, X., Zuo, X., Kucejova, B. & Chen, X. J. Reduced cytosolic protein synthesis suppresses mitochondrial degeneration. *Nature Cell Biol.* **10**, 1090–1097 (2008).
11. Liu, Y., Wang, X. & Chen, X. J. Misfolding of mutant adenine nucleotide translocase in yeast supports a novel mechanism of Ant1-induced muscle diseases. *Mol. Biol. Cell* **26**, 1985–1994 (2015).
12. Huber, A. *et al.* Sch9 regulates ribosome biogenesis via Stb3, Dot6 and Tod6 and the histone deacetylase complex RPD3L. *EMBO J.* **30**, 3052–3064 (2011).
13. Gilbert, W. V., Zhou, K., Butler, T. K. & Doudna, J. A. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science* **317**, 1224–1227 (2007).
14. Balagopal, V. & Parker, R. Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. *Curr. Opin. Cell Biol.* **21**, 403–408 (2009).
15. Kryndushkin, D. S., Smirnov, V. N., Ter-Avanesyan, M. D. & Kushnirov, V. V. Increased expression of Hsp40 chaperones, transcriptional factors, and ribosomal protein Rpp0 can cure yeast prions. *J. Biol. Chem.* **277**, 23702–23708 (2002).
16. Lee, A. S., Burdeinick-Kerr, R. & Whelan, S. P. A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs. *Proc. Natl Acad. Sci. USA* **110**, 324–329 (2013).
17. Xue, S. *et al.* RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation. *Nature* **517**, 33–38 (2015).
18. Begley, U. *et al.* Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Mol. Cell* **28**, 860–870 (2007).
19. Thorsness, P. E., White, K. H. & Fox, T. D. Inactivation of *YME1*, a member of the *ftsH-SEC18–PAS1-CDC48* family of putative ATPase-encoding genes, causes increased escape of DNA from mitochondria in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **13**, 5418–5426 (1993).
20. Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. & Pfanner, N. Importing mitochondrial proteins: machineries and mechanisms. *Cell* **138**, 628–644 (2009).
21. Dunn, C. D. & Jensen, R. E. Suppression of a defect in mitochondrial protein import identifies cytosolic proteins required for viability of yeast cells lacking mitochondrial DNA. *Genetics* **165**, 35–45 (2003).
22. Rikhvanov, E. G. *et al.* Do mitochondria regulate the heat-shock response in *Saccharomyces cerevisiae*? *Curr. Genet.* **48**, 44–59 (2005).
23. Chiabudini, M., Conz, C., Reckmann, F. & Rospert, S. Ribosome-associated complex and Ssb are required for translational repression induced by polylysine segments within nascent chains. *Mol. Cell. Biol.* **32**, 4769–4779 (2012).
24. Willmund, F. *et al.* The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. *Cell* **152**, 196–209 (2013).
25. Steffen, K. K. *et al.* Yeast life span extension by depletion of 60S ribosomal subunit is mediated by Gcn4. *Cell* **133**, 292–302 (2008).
26. Merkwirth, C. *et al.* Loss of prohibitin membrane scaffolds impairs mitochondrial architecture and leads to tau hyperphosphorylation and neurodegeneration. *PLoS Genet.* **8**, e1003021 (2012).
27. Segref, A. *et al.* Pathogenesis of human mitochondrial diseases is modulated by reduced activity of the ubiquitin/proteasome system. *Cell Metab.* **19**, 642–652 (2014).
28. Nargund, A. M., Pellegrino, M. W., Fiorese, C. J., Baker, B. M. & Haynes, C. M. Mitochondrial import efficiency of ATFS-1 regulates mitochondrial UPR activation. *Science* **337**, 587–590 (2012).
29. Johnson, S. C. *et al.* mTOR inhibition alleviates mitochondrial disease in a mouse model of Leigh syndrome. *Science* **342**, 1524–1528 (2013).

Author Contributions X.W. performed most of the experiments described in the study. X.J.C. conceived the project, guided the experiments, constructed yeast strains and performed genetic analysis. X.J.C. and X.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.J.C. (chenx@upstate.edu).

METHODS

No statistical methods were used to predetermine sample size, and the experiments were not randomized or blinded.

Media and strains. Yeast complete medium (YP) containing 1% Bacto yeast extract and 2% Bacto peptone is supplemented with 2% glucose (YPD) or 2% galactose plus raffinose (YPGal+Raf) or 2% glycerol (YPGly). Yeast minimal medium (YNBD) contains 0.67% Difco yeast nitrogen base without amino acids and 2% glucose. Supplements essential for auxotrophic strains were added to $20\,\mu g\,ml^{-1}$ for bases and amino acids except for leucine ($30\,\mu g\,ml^{-1}$) and lysine ($30\,\mu g\,ml^{-1}$). For ρ^{o}-lethality test, cells were spotted on YPD supplemented with ethidium bromide ($16\,\mu g\,ml^{-1}$). When necessary, cells were pre-grown on the ethidium bromide plates before being tested. Unless specified, all the yeast strains used in this study are isogenic to M2915-6A (*MAT*a, *ade2*, *ura3*, *leu2*). M2915-6A/α was derived from M2915-6A by mating-type switching. The null alleles for *DHH1*, *PBP1*, *RPL40A*, *BLM10*, *POC4*, *RPN4*, *SSB1*, *SSB2* and *UMP1*, marked by *kan*, were transferred from strains of the yeast knockout collection into the M2915-6A background by PCR amplification and one-step gene replacement. *YME1* was disrupted by the insertion of *LEU2*. PG1 (*MAT*α, *his3Δ1*, *leu2Δ0*, *lys2Δ0*, *ura3Δ0*, *trp1::GAL10-AAC2^{A128P}-HIS3*) was derived from BY4742 by replacing *TRP1* with the *GAL10-AAC2^{A128P}-HIS3* cassette by selecting His⁺ transformants. The strain used for iTRAQ analysis was CY4097 (*MAT*a, *ade2*, *ura3*, *leu2*, *trp1Δ::AAC2^{A128P}-URA3*, *lys2Δ::AAC2^{A128P}-kan*). To overexpress *SSB1*, the coding sequence together with its endogenous promoter and 3′ poly(A) tail was first cloned into the integrative pUC-URA3/4 vector. The resulting plasmid DNA was linearized by StuI within *URA3*, and then integrated into the *ura3* locus in M2915-6A. The Ura⁺ transformants were selected. The correct integration site and the copy number of the integrated *SSB1* were determined by Southern blot analysis.

Screen of anti-degenerative genes. A 2-μm-based yeast multicopy genomic library was transformed into the PG1 strain. Leu2⁺ transformants were replica-plated onto YPGal medium and 336 positive transformants were identified. The yeast plasmids containing the suppressor loci were rescued from yeast cells and amplified in *Escherichia coli*. The insert DNA in the suppressor plasmids was sequenced, subcloned, and confirmed by re-transforming into the PG1 strain followed by a phenotypic test.

Protein synthesis assay. Cells were grown in YPD at 30 °C overnight and then sub-cultured into synthetic complete medium lacking methionine at $A_{600\,nm}$ of ~0.3. Cells were then incubated at 25 °C for 3 h. Cell density was determined by measuring the culture at A_{600} nm. One millilitre of cell culture was labelled with $5.5\,\mu Ci$ [³⁵S]L-methionine/L-cysteine (EXPRE³⁵S³⁵S Protein Labelling Mix; >1,000 Ci mmol⁻¹; PerkinElmer NEG072) for 0, 5, 10, 20 and 30 min at 25 °C. Labelling was stopped by adding 1/10 volume of 100% TCA to each culture and chilling on ice. After heating at 90 °C for 20 min, protein precipitates were collected on GF/C filters (Whatman 1822-025), precipitated again with 10 ml of 10% TCA, washed with 10 ml of 95% ethanol, and then counted in 5 ml of UniverSol Liquid Scintillation Cocktail (MP Biomedical 882462). The count of each sample was normalized by its cell density.

Preparation of cytosolic proteins. Two independent replicates of the wild-type (M2915-6A) and the *AAC2^{A128P}* strain CY4097 were grown at 30 °C for 48 h in 50 ml of YPD and 100 ml of YNBD supplemented with leucine, lysine, tryptophan and adenine, respectively. The cells were then inoculated into 1 l of YPD and grown at 22 °C for 24 h. Cells were collected from the four independent cultures. A protease inhibitor cocktail free of EDTA and 50 μM MG132 were added to the spheroplasts before cell lysis by homogenization. Mitochondria were eliminated from the cytosol by differential centrifugation according to established procedures[30].

Mass spectrometry analysis. iTRAQ analysis was performed at Cornell Proteomics and Mass Spectrometry Facility as previously described[31]. The concentration of cytosolic proteins from two biological replicate samples was determined by Bradford assay using BSA as a calibrant, and further quantified by running on a precast NOVEX 12% Tris/Glycine mini-gel (Invitrogen) along with a series of *E. coli* lysates (2, 5, 10, 20 μg per lane). The SDS gel was visualized with colloidal Coomassie blue stain (Invitrogen), and imaged by Typhoon 9400 scanner followed by ImageQuant Software version 5.2 (GE Healthcare). An aliquot (100 μg) of proteins in a total volume of 20 μl was denatured by adding 1 μl of 2% SDS and reduced with 2 μl of 50 mM Tris-(2-carboxyethyl)phosphine (TCEP). Cysteine residues were blocked with 1 μl of 200 mM methyl methanethiosulfonate (MMTS) using the iTRAQ Reagents kit (AB Sciex). The proteins were then precipitated with acetone and digested with trypsin. Peptides were labelled with iTRAQ reagents per manufacturer's instructions (AB Sciex). The 114- and 115-tags were used to label the peptides from the control strain (M2915-6A), while the 116- and 117-tags were used to label those from the *AAC2^{A128P}*-expressing cells. The four samples were then pooled and subjected to fractionation using high pH

reverse phase liquid chromatography onto an XTerra MS C18 column (3.5 μm, 2.1 × 150 mm) from Waters as previously reported previously[32]. Twelve peptide fractions were then analysed by nanoLC-MS/MS on a LTQ-Orbitrap Velos (Thermo-Fisher Scientific) mass spectrometer equipped with nano ion source using high energy collision dissociation (HCD) as previously reported[32]. The data files acquired were converted into MGF files using Proteome Discoverer 1.3 (PD 1.3, Thermo). Subsequent database search were carried out by Mascot Daemon (version 2.3, Matrix Science) against the uni-protein database of *Saccharomyces cerevisiae* (5,882 entries downloaded on 12 Mar 2012) for both protein identification and iTRAQ quantitation. The default Mascot search settings were as follows: one missed cleavage site by trypsin allowed with fixed MMTS modification of cysteine, fixed four-plex iTRAQ modifications on Lys and amino-terminal amines and variable modifications of methionine oxidation, deamidation of Asn and Gln residues, and 4-plex iTRAQ on Tyr. The peptide and fragment mass tolerance values were 20 p.p.m. and 0.1 Da, respectively. Only peptides with significant scores at the 99% confidence interval were counted as identified. For quantitation, we used the requirement of two unique peptide hits for each protein. The quantitative protein ratios were weighted and normalized by the median ratio for each set of experiments. The threshold of significance between the proteins from mutant and wild-type cells was further accessed as previously described[31,33,34]. The internal error between the biological replicates, defined as the absolute value of the difference between $\log_2(116/114)$ and $\log_2(117/115)$ at which 95% of all proteins had no deviation from each other, was calculated to be 0.34 at $\log_2$ scale. Relative ratios of ± 1.266 were thus deduced as significant difference with a statistical confidence of <0.05.

Ssb1–TAP pull-down. The wild-type strain CY5371 (*MAT*a, *met15Δ0*, *his3Δ0*, *leu2Δ0*, *ura3Δ0*, *SSB1-TAP::HIS3*) and the isogenic *AAC2^{A128P}*-expressing strain CY5171 (*MAT*a, *met15Δ0*, *his3Δ0*, *leu2Δ0*, *SSB1-TAP::HIS3*, *ura3Δ:: AAC2^{A128P}::URA3*) were grown in 100 ml of YPD at 22 °C for 48 h. The cells were then inoculated into two litres of YPD, and grown at 22 °C for 24 h. Cells were collected by centrifugation, resuspended to 1 g ml⁻¹ with buffer Z (20 mM Tris-HCl, pH 8, 150 mM KCl, 1 mM EDTA, pH 8, 10% glycerol, 1 mM dithiothreitol (DTT), 1 mM phenylmethanesulfonylfluoride (PMSF) and protease cocktail), and then passed through microfluidizer 10 times at 18,000 p.s.i. Tween-20 was added to cell lysates to the concentration of 0.1%. Cell lysates were shaken at 4 °C for 20 min. Cell debris was removed by spinning at 12,000g at 4 °C for 10 min. Cell lysates were further purified by centrifugation at 100,000g (Beckman 50.2 Ti rotor) at 4 °C for 100 min, and then bound to buffer Z (+0.1% Tween-20) equilibrated rabbit IgG-Agarose (Sigma A2909) by incubation at 4 °C for 3 h. After binding, the rabbit IgG-agarose beads were washed five times with buffer Z and 0.1% Tween-20 to remove the unbound contaminants. TAP-tagged Ssb1 and its associated proteins were eluted by incubating with buffer Z plus 0.1% Tween-20 and 100 U of ProTEV Plus (Promega V610A) at 4 °C overnight. The eluates were then subjected to calmodulin binding in buffer Z with 0.1% Tween-20, 150 mM NaCl, 1.5 mM magnesium acetate, 1 mM imidizole plus 5 mM CaCl₂ at 4 °C overnight. The beads were then washed five times with the above binding buffer to remove the unbound proteins. The TAP-tagged Ssb1 and its associated proteins were eluted by incubating with buffer Z, 0.1% Tween-20, 1.5 mM magnesium acetate, 1 mM imidizole and 2.5 mM EGTA at 4 °C.

Haemagglutinin-tagging and protein stability measurement. A 3×HA tag was added to the carboxy terminus of Gis2 and Nog2 by integration of a PCR product amplified from a 3×HA–KanMX cassette. The primers used for tagging *GIS2* were GIS2-3HA-KAN-ITFP (5′-taacgaaactggccatatttccaaggattgtccaaaggctCGG ATCCCCGGGGTTAATTAA-3′) and GIS2-3HA-KAN-ITRP (5′-taagtgattgttcct ttgtttttttcagtggatgtttcacGAATTCGAGCTCGTTTAAAC-3′). The *GIS2* sequence is shown in lowercase, whereas the sequence complementary to the 3×HA– KanMX6 cassette is in uppercase. The primers used for tagging *NOG2* were NOG2-3HA-KAN-ITFP (5′-ggaagaaaaaccaaagaagaaagaagttgagaagacggcaCGGA TCCCCGGGGTTAATTAA-3′) and NOG2-3HA-KAN-ITRP (5′-tcttaaattaatattat acaccggttgtccgttttttacctGAATTCGAGCTCGTTTAAAC-3′). The *NOG2* sequence is shown in lowercase, whereas the sequence complementary to the 3×HA– KanMX6 cassette is in uppercase. To detect the steady state levels of the HA-tagged proteins, the cells were grown at 25 °C until all the strains reached $A_{600\,nm}$ = 3.5–5.0. Cells equivalent to $A_{600\,nm}$ values of 3 were collected and subjected to protein extraction. Cell lysates were analysed by SDS–PAGE and levels of HA-tagged proteins were determined by western blot using anti-HA antibody (Covance MMS-101R). Relative protein levels were quantified by ImageJ. To measure protein stability, cycloheximide was added to the cultures at a concentration of 200 μg ml⁻¹, and cells were collected after incubation at 25 °C for 0, 10, 20, 40, 60, 80 and 120 min.

Co-immunoprecipitation and ubiquitination assay of Gis2. M2915-6A (wild type), UPU5-7C (*AAC2^{A128P}*), CY5738 (*GIS2-HA*) and CY5740 (*GIS2-HA*, *AAC2^{A128P}*) were grown in YPD to an $A_{600\,nm}$ of 3.0. Cells equivalent to an

$A_{600\,nm}$ of 20 were collected and resuspended in 0.4 ml of lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 25 mM N-ethylmaleimide, 5 mM 2-mercarptoethanol, 50 μM MG132, and complete protease inhibitor cocktail (Roche 11697498001). Cells were lysed with 0.4 ml of glass beads by using BIOSPEC Mini-Beadbeater-16. The total protein concentration was determined by Quick Start Bradford Protein Assay (Bio-Rad 500-0201). Equal amounts of cell lysates were bound to 50 μl 50% slurry of monoclonal anti-HA agarose conjugate clone HA-7 (Sigma: A2095) at 4 °C for 2 h. The resin was precipitated by centrifugation and washed five times with lysis buffer. The proteins were eluted with 25 μl 2×SDS sample buffer (125 mM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 5% 2-mercaptoethanol, 0.005% bromophenol blue), analysed by SDS–PAGE and probed with anti-HA (Cell Signaling 3724S) and anti-ubiquitin (Thermo Scientific 701339) antibodies.

Ribosomal profile analysis. M2915-6A (WT) and CY3555 ($trp1\Delta::AAC2^{A128P}$-kan) were transformed with either the multicopy vector pRS426 or pRS426-Nog2. The strains were grown in 100 ml of YNBD plus casamino acids supplemented with adenine and tryptophan to an $A_{600\,nm}$ of about 0.8. The culture was added with 0.1 mg ml^{-1} cycloheximide, followed by a rapid swirl and incubation on ice for 10 min. The cells were collected by contrifugation at 4 °C and washed with lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 30 mM MgCl$_2$, 100 μg ml^{-1} cycloheximide, 200 μg ml^{-1} heparin and 0.1% diethylpyrocarbonate). The cells were then resuspended in 1 ml of lysis buffer containing 0.4 ml of acid-washed, sterile and cold glass beads. The cells were disrupted by BIOSPEC Mini-Beadbeater-16 ten times for a 30-s beating with a 60-s incubation on ice. The cell lysate was subjected to centrifugation at 5,000 g at 4 °C for 5 min and 160,000g at 4 °C for 10 min. A portion of supernatant with 6 units of absorbance at 260 nm was added to the top of a 12-ml 7–47% (w/v) continuous sucrose gradient containing 50 mM Tris-HCl, pH 7.4, 50 mM NH$_4$Cl, 12 mM MgCl$_2$ and 1 mM DTT. The sucrose gradient was centrifuged at 280,000g at 4 °C for 2.5 h in a Beckman SW40Ti rotor. The ribosome profile was analysed with the Density Gradient Fractionation System (Brandel BR-188-176).

30. Diekert, K., de Kroon, A. I. P. M., Kispal, G. & Lill, R. in *Mitochondrion* (eds Pon, L. A. & Schon, E. A.) 41–44 (Academic, 2001).
31. Yang, Y. *et al.* Evaluation of different multidimensional LC-MS/MS pipelines for isobaric tags for relative and absolute quantitation (iTRAQ)-based proteomic analysis of potato tubers in response to cold storage. *J. Proteome Res.* **10,** 4647–4660 (2011).
32. Yang, Q. S. *et al.* Quantitative proteomic analysis reveals that antioxidation mechanisms contribute to cold tolerance in plantain (*Musa paradisiaca* L.; ABB Group) seedlings. *Mol. Cell. Proteomics* **11,** 1853–1869 (2012).
33. Redding, A. M., Mukhopadhyay, A., Joyner, D. C., Hazen, T. C. & Keasling, J. D. Study of nitrate stress in *Desulfovibrio vulgaris* Hildenborough using iTRAQ proteomics. *Brief. Funct. Genomic Proteomic* **5,** 133–143 (2006).
34. Gan, C. S., Chong, P. K., Pham, T. K. & Wright, P. C. Technical, experimental and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J. Proteome Res.* **6,** 821–827 (2007).
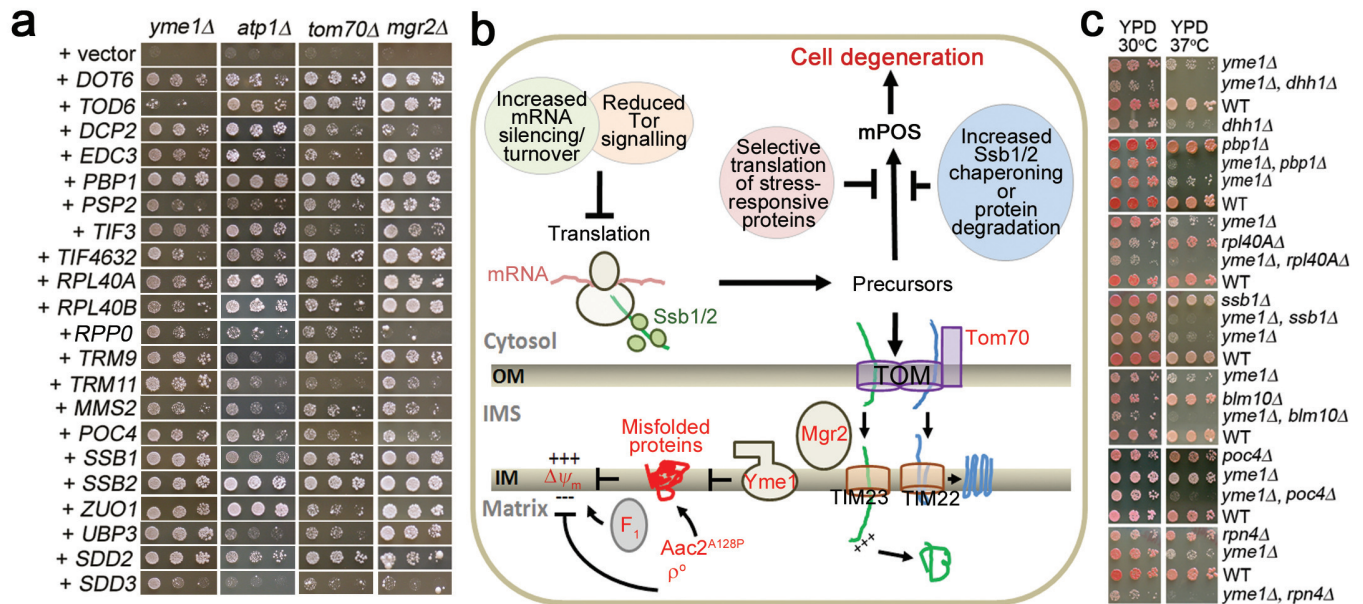
Extended Data Figure 1 | Proteostatic crosstalk between mitochondria and the cytosol. a, Suppression of ρ°-lethality in the *yme1Δ*, *atp1Δ*, *tom70Δ* and *mgr2Δ* mutants by anti-degenerative genes on YPD medium supplemented with ethidium bromide that eliminates mtDNA. **b**, Schematic depiction of the cytosolic anti-degenerative proteostatic network that suppresses mPOS.

c, Synthetic growth defect between *yme1Δ* and the disruption of anti-degenerative genes (*PBP1*, *RPL40A*, *SSB1* and *POC4*) and other genes affecting cytosolic proteostasis (*DHH1*, *BLM10* and *RPN4*). Cells were grown on YPD medium and incubated at the indicated temperatures.
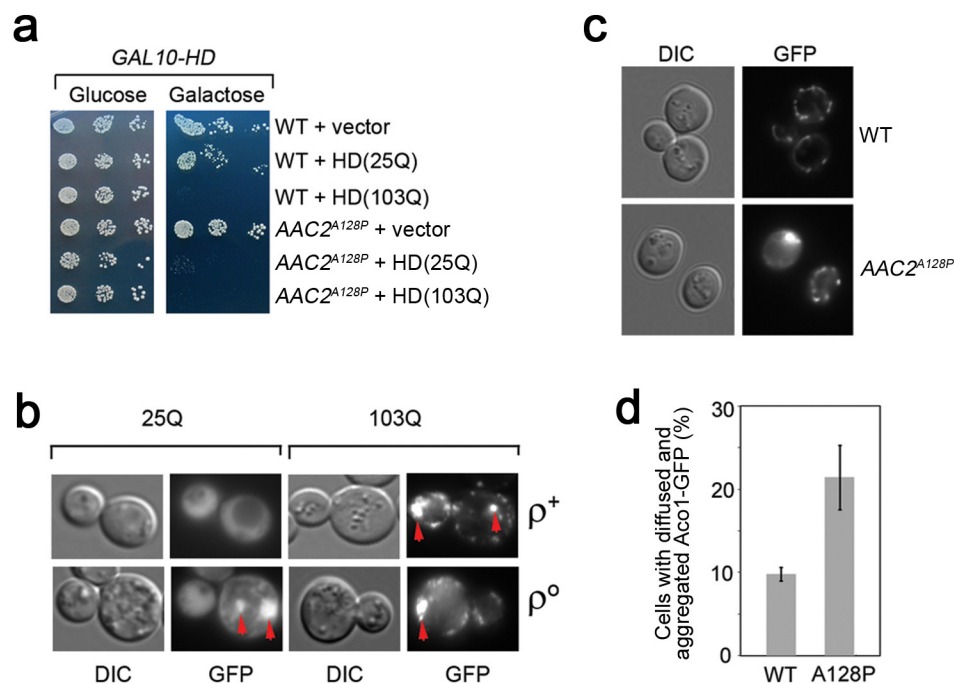
Extended Data Figure 2 | Mitochondrial damage increases protein aggregation in the cytosol. **a**, Synthetic lethality between mitochondrial damage and cytosolic protein misfolding. Growth of $AAC2^{A128P}$ but not wild-type cells is inhibited by expression of $GAL10$-HD(25Q) on galactose medium, whereas growth of both $AAC2^{A128P}$ and wild-type cells is inhibited by expression of $GAL10$-HD(103Q). Yeast transformants were serially diluted in water and spotted on minimal glucose or galactose medium. The plates were incubated at 25 °C for 4 days. **b**, HD(25Q) forms aggregates in ρ^o but not ρ^+ cells, whereas HD(103Q) is aggregated in both types of cells. The images are representatives of 200 cells examined for each strain. **c**, Increased cytosolic accumulation and aggregation of green fluorescent protein (GFP)-tagged Aco1 in $AAC2^{A128P}$-expressing cells. Representative images are from four independent experiments, with a total number of 2,397 and 1,608 cells examined for the wild-type and $AAC2^{A128P}$ strains, respectively. **d**, Mean ± s.d. of the four experiments in **c** ($P < 0.001$, unpaired Student's t-test).
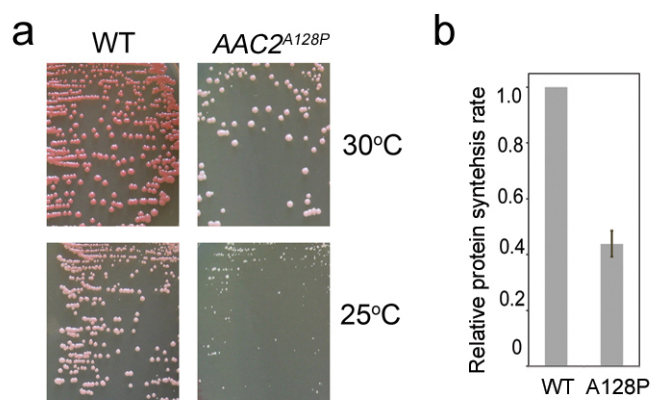
Extended Data Figure 3 | Growth phenotype and relative protein synthesis rate of yeast cells expressing *AAC2^A128P*. **a**, Cells were grown on YPD medium at indicated temperatures for four days. **b**, Expression of *AAC2^A128P* reduces the global protein synthesis rate, which is measured by the incorporation of [^{35}S]methionine after incubating at 25 °C for 5 min. Data are mean ± s.d. of three independent experiments ($P < 0.005$, unpaired Student's *t*-test).
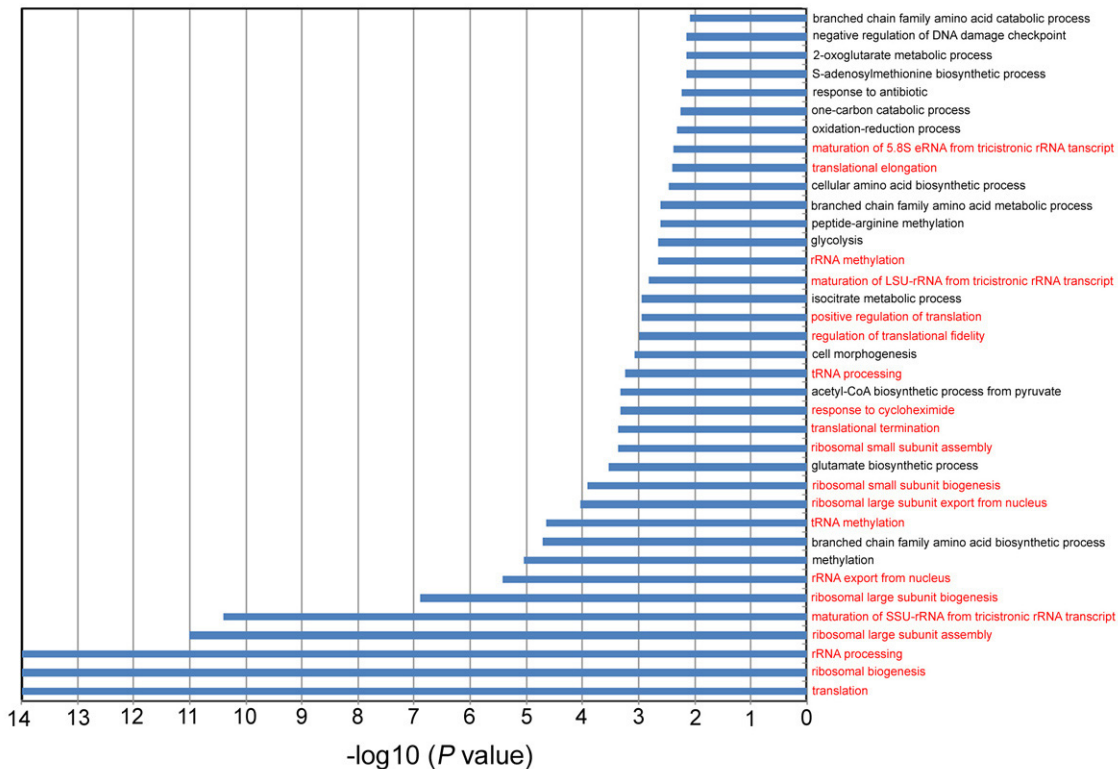
Extended Data Figure 4 | Gene Ontology analysis of proteins that are upregulated in the cytosol of *AAC2*[A128P] cells. Coloured in red are those involved in ribosomal biogenesis/translation.
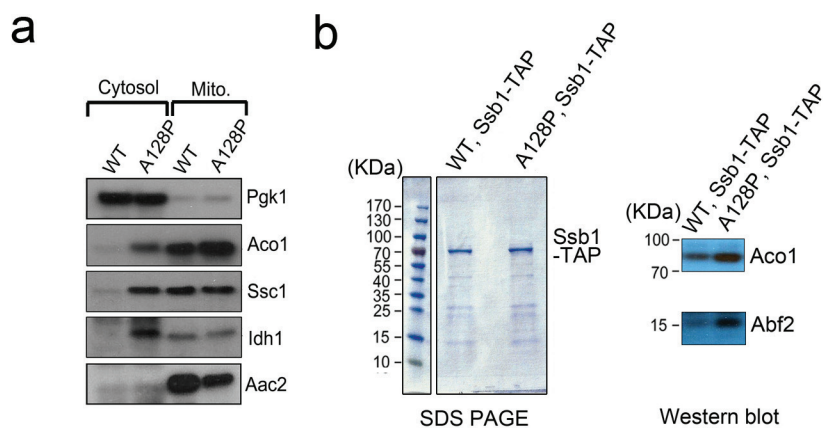
Extended Data Figure 5 | Increased cytosolic retention and Ssb1 association of mitochondrial precursors in *AAC2^{A128P}* cells. a, Western blot showing increased retention of representative mitochondrial proteins (Aco1, Ssc1, Idh1 and Aac2) in the cytosol of *AAC2^{A128P}* cells. Pgk1, the cytosolic 3-phosphoglycerate kinase, is used as a control. **b**, SDS–PAGE (left) and western blot (right) showing Coomassie-stained Ssb1–TAP (tandem affinity purification) pull-down products and increased association of Ssb1–TAP with the mitochondrial Aco1 and Abf2 in *AAC2^{A128P}* cells, respectively. Full scans of blots and gels are available in the Supplementary Information.
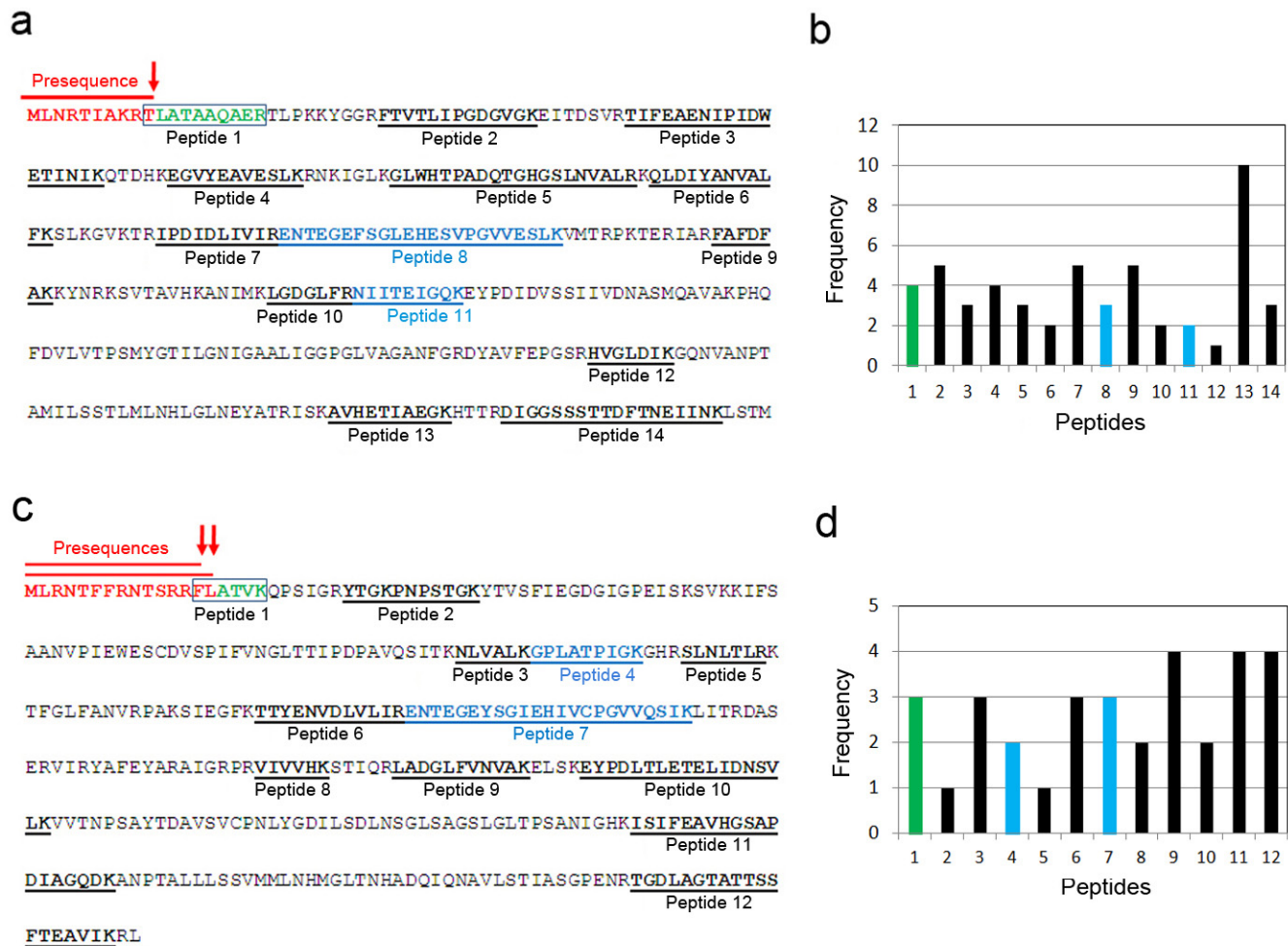
Extended Data Figure 6 | iTRAQ analysis showing the presence of Idh1 and Idh2 precursors in the cytosol of *AAC2^A128P*-expressing cells. a, Underlined are Idh1 peptides detected by mass spectrometry. The presequence of Idh1 is shown in red. The cleavage site of mitochondrial peptidase for Idh1 maturation is indicated by the red arrow. Boxed is the peptide 1 detected by mass spectrometry after trypsinization. Peptide 1 encompasses the last threonine residue of the presequence, suggesting that it is derived from the Idh1 precursor instead of its mature form. **b**, Frequency of Idh1 peptides identified by mass spectrometry. **c**, Underlined are Idh2 peptides detected by mass spectrometry. The presequences of Idh2 are shown in red. The cleavage sites of mitochondrial peptidase for Idh2 maturation are indicated by the red arrows. Boxed is the peptide 1 detected by mass spectrometry after trypsinization. Peptide 1 encompasses the last 1–2 residues of the presequence, suggesting that it is derived from the Idh2 precursor instead of its mature form. **d**, Frequency of Idh2 peptides identified by mass spectrometry.
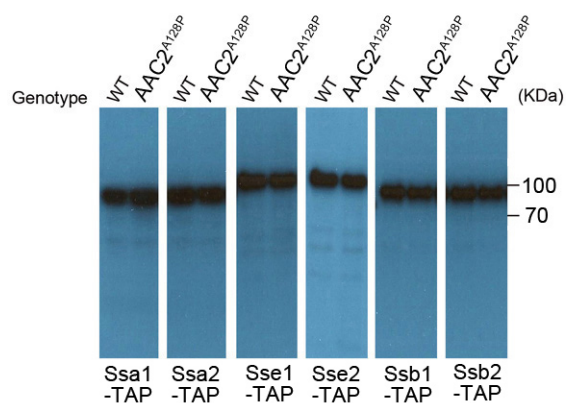
Extended Data Figure 7 | Western blot showing the expression levels of TAP-tagged Ssa1, Ssa2, Sse1, Sse2, Ssb1 and Ssb2 in wild-type and *AAC2^{A128P}* cells. Equal amounts of lysates from cells grown at 25 °C were analysed using an antibody against protein A in the TAP tag. Full scans of blots are available in the Supplementary Information.
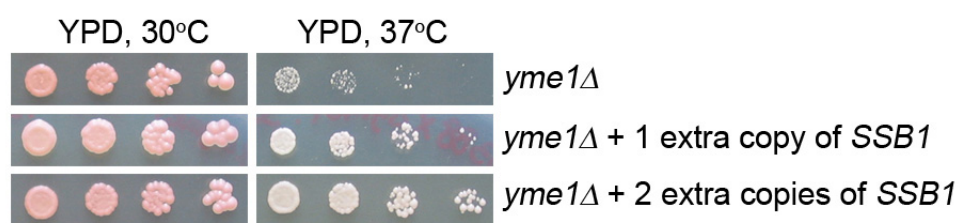
Extended Data Figure 8 | The heat sensitivity of the *yme1Δ* mutant is suppressed by one and two extra copies of *SSB1* integrated into the genome. Cells were diluted in water, spotted on YPD plates and incubated at the indicated temperatures for 3 days.
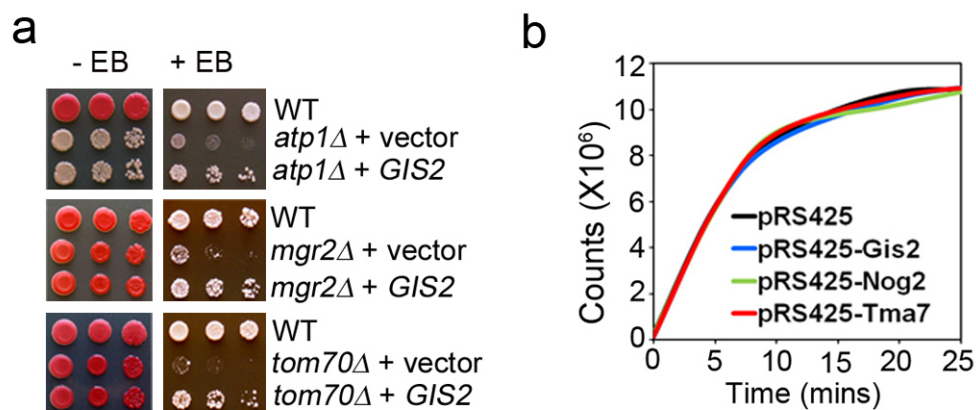
Extended Data Figure 9 | Suppression of ρ°-lethality in *atp1Δ, mgr2Δ* and *tom70Δ* mutants by overexpression of *GIS2* and protein synthesis rate in cells overexpressing *GIS2, NOG2* and *TMA7*. **a**, Ethidium bromide (EB) sensitivity test. Yeast transformants were diluted in water and spotted on YPD with or without ethidium bromide. The plates were incubated at 30 °C for four days. *GIS2* was overexpressed from the multicopy vector pRS425. **b**, *In vivo* protein synthesis assay. The incorporation of [³⁵S]methionine in the wild-type cells overexpressing *GIS2, NOG2* and *TMA7* on a multicopy vector was measured at 25 °C.
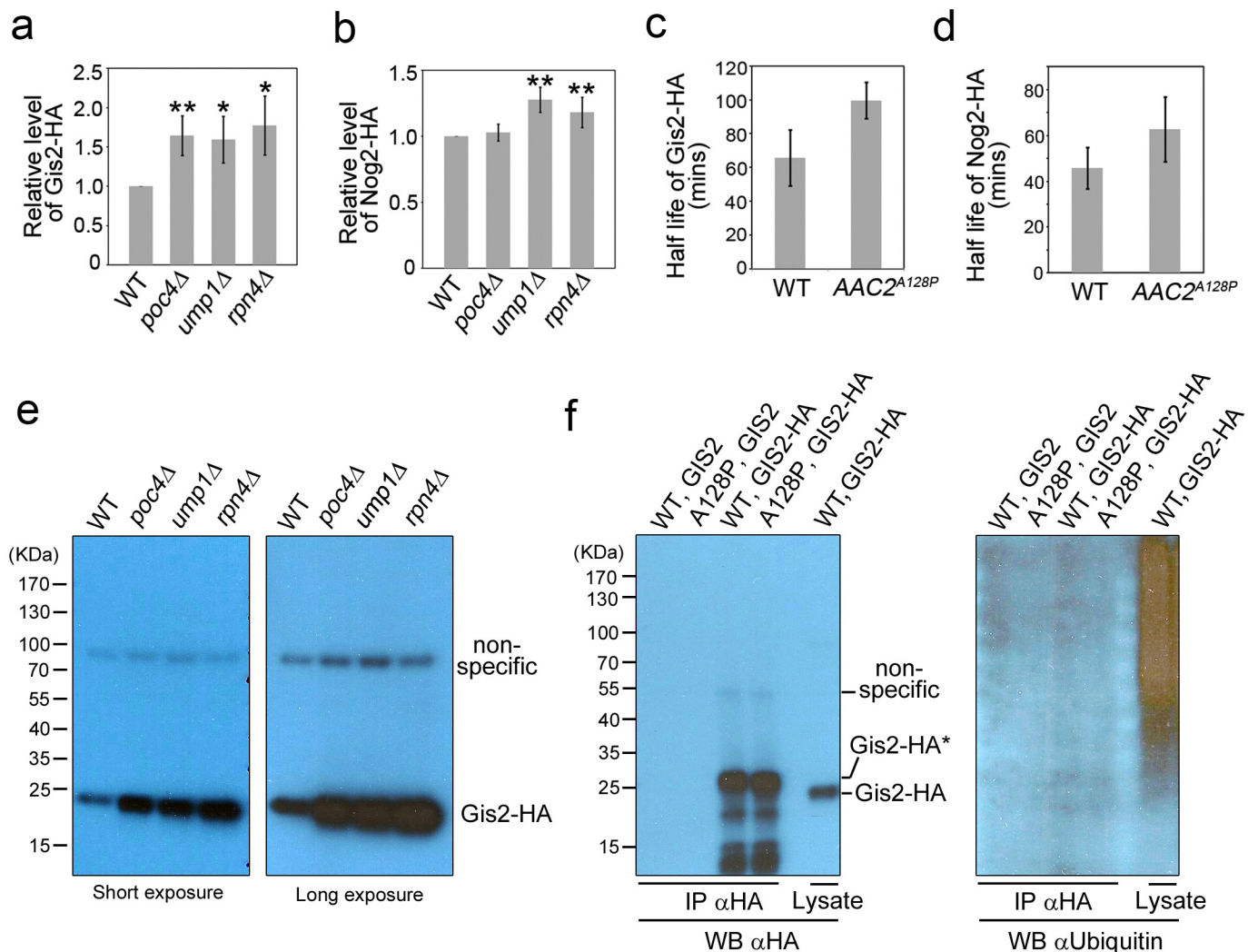
Extended Data Figure 10 | Stability of Gis2–HA and Nog2–HA.
a, **b**, Relative steady-state levels of Gis2–HA and Nog2–HA in proteasomal mutants. Data are mean ± s.d. of 3 and 5 independent experiments for Gis2–HA and Nog2–HA, respectively (*$P < 0.05$; **$P < 0.01$; unpaired Student's t-test). **c**, **d**, Half-life of Gis2–HA and Nog2–HA in the wild-type and $AAC2^{A128P}$ (A128P) cells. Data are mean ± s.d. of three experiments ($P < 0.05$ for Gis2–HA and $P = 0.15$ for Nog2–HA, unpaired Student's t-test). **e**, Western blot analysis showing no evidence of Gis2–HA ubiquitination in the $poc4\Delta$, $ump1\Delta$ and $rpn4\Delta$ cells. The $ump1\Delta$ and $rpn4\Delta$ cells are temperature-sensitive because of defective proteasomal function. Cells were grown at the non-permissive temperature (37 °C) before being lysed for protein extraction and SDS–PAGE. **f**, Gis2–HA was immunoprecipitated from the wild-type (WT) and $AAC2^{A128P}$ (A128P) cells and analysed by western blot using antibodies against HA (left) and ubiquitin (right). Note that the immunoprecipitation-purified full-length Gis2–HA migrates slower than the protein in the cell lysate, probably due to posttranslational modification during immunoprecipitation. The cryptic modification is unrelated to ubiquitination, based on the lack of reactivity with the anti-ubiquitin antibody.

LETTER

Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol

Lidia Wrobel[1]*, Ulrike Topf[1]*, Piotr Bragoszewski[1], Sebastian Wiese[2,3]†, Malgorzata E. Sztolsztener[1], Silke Oeljeklaus[2], Aksana Varabyova[1], Maciej Lirski[4], Piotr Chroscicki[1], Seweryn Mroczek[4,5], Elzbieta Januszewicz[1], Andrzej Dziembowski[4,5], Marta Koblowska[4,6], Bettina Warscheid[2,3] & Agnieszka Chacinska[1]

Most of the mitochondrial proteome originates from nuclear genes and is transported into the mitochondria after synthesis in the cytosol. Complex machineries which maintain the specificity of protein import and sorting include the TIM23 translocase responsible for the transfer of precursor proteins into the matrix, and the mitochondrial intermembrane space import and assembly (MIA) machinery required for the biogenesis of intermembrane space proteins. Dysfunction of mitochondrial protein sorting pathways results in diminishing specific substrate proteins, followed by systemic pathology of the organelle and organismal death[1–4]. The cellular responses caused by accumulation of mitochondrial precursor proteins in the cytosol are mainly unknown. Here we present a comprehensive picture of the changes in the cellular transcriptome and proteome in response to a mitochondrial import defect and precursor over-accumulation stress. Pathways were identified that protect the cell against mitochondrial biogenesis defects by inhibiting protein synthesis and by activation of the proteasome, a major machine for cellular protein clearance. Proteasomal activity is modulated in proportion to the quantity of mislocalized mitochondrial precursor proteins in the cytosol. We propose that this type of unfolded protein response activated by mistargeting of proteins (UPRam) is beneficial for the cells. UPRam provides a means for buffering the consequences of physiological slowdown in mitochondrial protein import and for counteracting pathologies that are caused or contributed by mitochondrial dysfunction.

The MIA machinery constitutes one of the essential protein translocation pathways into mitochondria and is responsible for the import of proteins targeted to the intermembrane space (IMS) of mitochondria[1–4]. RNA sequencing (RNA-seq) analysis of *mia40-4int*[5,6] (a yeast temperature-sensitive mutant defective for protein import into the IMS) compared with wild-type cells was performed under permissive growth conditions to avoid cumulative consequences of broad mitochondrial dysfunction (Supplementary Table 1). Transcripts for cytosolic ribosomal proteins and proteins involved in translation, but not mitochondrial ribosomal proteins, decreased (Fig. 1a and Extended Data Fig. 1a–c). Thus, changes in the transcriptome could not be directly translated into protein abundance. Next, the cellular proteomes of *mia40-4int*[S] and wild-type[S] cells were determined under permissive growth conditions using stable isotope labelling by amino acids in cell culture (SILAC)[7] (Supplementary Table 2 and Extended Data Fig. 1d–g). Sensitive MIA substrates were significantly decreased in the *mia40-4int*[S] proteome (Supplementary Table 3 and Extended Data Fig. 1h–k) because they were not maintained upon mitochondrial import inhibition[8].

We determined the relative abundance of 2,564 proteins. A small fraction (11%) exhibited significant changes in abundance with 112 downregulated and 174 upregulated proteins in *mia40-4int*[S] (Supplementary Table 3 and Extended Data Fig. 1k). Localization and Gene Ontology (GO) analysis revealed the modulation of the structure and function of genetic information and reshaping of metabolic networks in *mia40-4int*[S] (Extended Data Fig. 2a, b). Consistent with the RNA-seq results (Fig. 1a), a class of proteins that constitute the cytosolic ribosome were downregulated (Fig. 1b and Extended Data Fig. 2b, c). Under restrictive growth conditions an imbalance in the ratio between 25S and 18S rRNA (Fig. 1c, d) and a change in ribosome content with less prominent polysomes in the *mia40-4int* mutants (Extended Data Fig. 2d–f) were observed. Interestingly, the *mia40-4int* mutant exhibited a decrease in protein synthesis (Fig. 1e), but extensive cellular death was excluded as a reason for this decrease (Extended Data Fig. 3a, b). The *mia40-4int* cells transformed with wild-type *MIA40* exhibited a normal protein synthesis rate demonstrating that the translation defect depends on *mia40-4int* (Fig. 1e and Extended Data Fig. 3c).

The ubiquitin–proteasome system (UPS)[9–12] is involved in the degradation of MIA substrate proteins[8,13]. Abundance of many proteins involved in the UPS did not change in *mia40-4int*[S], including the core proteasomal subunits (Supplementary Table 3 and Extended Data Fig. 2d). Components of the proteasome assembly chaperone complex, Irc25 (also known as Poc3) and Poc4 (refs 14–17), significantly increased at the protein level in *mia40-4int*[S] (Fig. 1f) despite non-significant changes at the mRNA level (Extended Data Fig. 3d). Other assembly chaperones, Ump1, Pba1 (also known as Poc1) and Add66 (also known as Poc2 or Pba2) (refs 14–17), showed a tendency to increase, whereas the stress-inducible proteasome chaperone Adc17 (also known as Tma17) (ref. 18) did not change (Supplementary Table 3). Under restrictive growth conditions, ubiquitinated protein species decreased in *mia40-4int* and the effect was complemented by *MIA40* (Fig. 1g). The abundance of ubiquitinated species may mirror translation efficiency because newly synthesized proteins constitute a substantial fraction of polyubiquitinated proteins in the cell[19]. However, the proteasomal inhibitor MG132 attenuated the difference in ubiquitination between wild-type and mutant cells, suggesting acceleration in *mia40-4* protein turnover (Fig. 1g). The *mia40-4int* cells exhibited higher proteasomal activity than wild-type cells and this effect was rescued by *MIA40* (Fig. 1h and Extended Data Fig. 4a–c). Applying restrictive temperatures to wild-type cells did not stimulate proteasomal activity (Extended Data Fig. 4d). Furthermore, other plasmid-borne conditional *mia40* mutants[5,6] with a phenotype milder than *mia40-4int* showed both a clear decrease in ubiquitinated species and an increase in proteasomal activity (Extended Data Fig. 4e, f).

Overexpression of Mia40 improved protein import into the mitochondrial IMS in human cells[20] and in yeast as determined by an
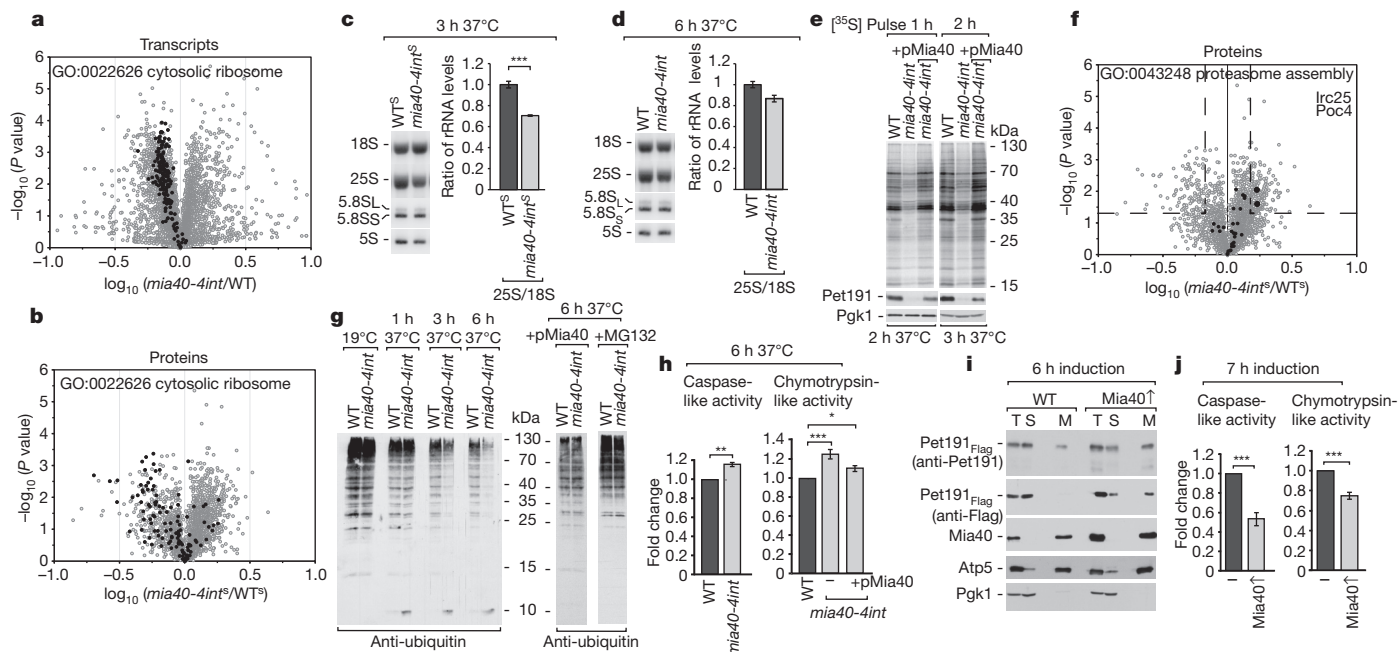
Figure 1 | The MIA pathway import efficiency regulates protein synthesis and proteasomal activity. a, b, Distribution of transcripts and proteins based on the GO term (black dots). c, d, Northern blot analysis of RNA and ratio of 25S/18S rRNAs in yeast grown in respiratory medium. Mean ± s.e.m., $n = 3$. e, Protein synthesis in cells grown in glycerol. f, Distribution of proteins based on GO terms (black dots). Dashed lines separate proteins of P value <0.05 and fold-change <−1.5 and >1.5. g, Ubiquitination analysis of cellular proteins.

h, Proteasomal activity. Mean ± s.e.m., $n = 7$ (caspase-like activity), $n = 4$ (chymotrypsin-like activity). i, Cells were fractionated into total (T), post-mitochondrial supernatant (S) and mitochondria (M). Mia40↑, Mia40 overproducing strain. j, Proteasomal activity in cells grown at 24 °C. Mean ± s.e.m., $n = 4$ (caspase-like activity), $n = 6$ (chymotrypsin-like activity). c, h, j, *$P < 0.05$; **$P < 0.03$; ***$P < 0.01$. WT, wild-type. c–e, g, i, Uncropped blots are in Supplementary Information Fig. 1.

in organello assay (Extended Fig. 5a–c) but did not influence the abundance of mitochondrial proteins (Extended Data Fig. 5d). However, when the MIA pathway was challenged by the overproduced Flag fusion protein Pet191$_{Flag}$[8], this protein was more efficiently localized to mitochondria (Fig. 1i and Extended Data Fig. 5e, f). The more productive removal of precursors from the cytosol by mitochondria with elevated Mia40 levels resulted in decreased proteasomal activity (Fig. 1j).

The import of a large fraction of mitochondrial proteins relies on the electrochemical potential of the inner mitochondrial membrane (IM

potential)[1–4]. Cells treated with the chemical uncoupler carbonyl cyanide *m*-chlorophenyl hydrazine (CCCP) to dissipate the IM potential accumulated non-imported mitochondrial proteins in their longer unprocessed precursor forms (with the mitochondrial targeting sequence, called presequences[1–4]). During the chase in the absence of CCCP, precursor proteins diminished (Fig. 2a and Extended Data Fig. 6a, b). Following the accumulation of mitochondrial precursors in the cytosol, a significant stimulation of the proteasome was observed (Fig. 2b). Similar accumulation of precursor proteins was observed in temperature-sensitive mutants of the TIM23 translocase and motor
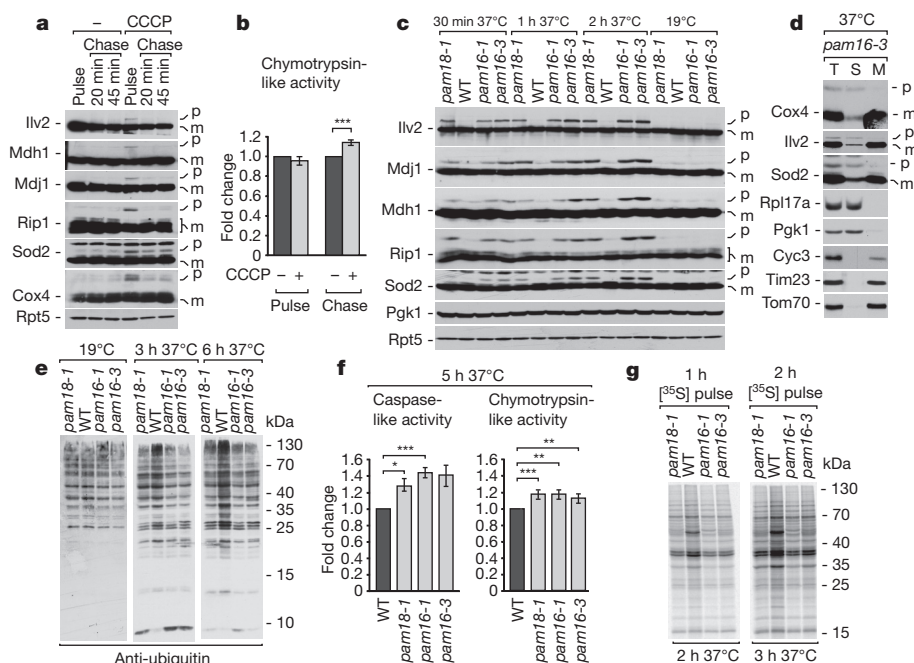
Figure 2 | Defects in the presequence import pathway modulate protein synthesis and proteasomal activity. a, b, Cells were treated for 30 min with CCCP (pulse) and chased without CCCP and analysed for protein content and proteasomal activity. Mean ± s.e.m., $n = 6$. c, The *pam18*, *pam16* and wild-type were analysed for protein content. d, The *pam16-3* cells after a 90 min shift to 37 °C were fractionated. Total (T); post-mitochondrial supernatant (S); mitochondria (M); precursor (p); mature (m). e, Cells were analysed for ubiquitination. f, Proteasomal activity. Mean ± s.e.m., $n = 3$ (caspase-like activity), $n = 5$ (chymotrypsin-like activity). g, Protein synthesis in cells grown in glycerol. b, f, *$P < 0.05$; **$P < 0.03$; ***$P < 0.01$. WT, wild-type. a, c–e, g, Uncropped blots are in Supplementary Information Fig. 1.

complex, *pam 18*, *pam16* and *tim17* (refs 21, 22) (Fig. 2c and Extended Data Fig. 6c). These precursor forms were found in the cytosol (Fig. 2d and Extended Data Fig. 6d). The *pam18* and *pam16* mutants did not show any major changes in protein abundance (Extended Data Fig. 6e). However, similar to *mia40*, the *pam18*, *pam16* and *tim17* mutants exhibited a clear decrease in ubiquitinated species and increase in proteasomal activity (Fig. 2e, f and Extended Data Fig. 6f–h). Although we did not detect any change in ribosomal proteins or ribosomal RNA content (Extended Data Fig. 6e, i), protein synthesis was decreased and was accompanied by an increase in monosomes in *pam16* cells (Fig. 2g and Extended Data Fig. 6j, k). Thus, the defective import of proteins into mitochondria leads to a response that regulates the cellular protein homeostasis through two arms, the attenuation of the cytosolic synthesis of proteins and activation of the proteasome to prevent the accumulation of mistargeted proteins.

To determine whether mistargeted mitochondrial proteins are responsible for proteasomal regulation, we used Pet191$_{Flag}$ and Mix17$_{Flag}$, which were partly mislocalized to the cytosol when ectopically expressed[8] but remained mainly soluble (Fig. 1i and Extended Data Figs 5f and 7a, b). Proteins that formed aggregates were previously shown to inhibit the proteasome[23]. The steady-state levels of cellular proteins, including ribosomal and proteasomal subunits, rRNA levels and cytosolic translation were unchanged upon the Pet191$_{Flag}$ and Mix17$_{Flag}$ overexpression (Extended Data Fig. 7c–e). However, in the presence of mistargeted Pet191$_{Flag}$ or Mix17$_{Flag}$ proteasomal activity increased substantially (Fig. 3a and Extended Data Fig. 7f). Thus, mitochondrial protein mislocalization uncoupled protein degradation and protein synthesis effects.

Other mitochondrial proteins, Cox4 and Mdj1 (substrates of the TIM23 pathway) and their non-imported variants that lacked presequences (mCox4 and mMdj1), were overexpressed (Extended Data Fig. 7g–i). The overexpression of Cox4 but not Mdj1, led to a visible accumulation of a cytosolic precursor form (Extended Data Fig. 7g, h). Cox4, mCox4, and mMdj1 (despite its low abundance) stimulated the

proteasome, in contrast to fully imported Mdj1 which did not stimulate the proteasome (Fig. 3b). Non-mitochondrial model proteins, including mouse dihydrofolate reductase DHFR and its conformationally destabilized variant DHFRds, temperature sensitive Ubc9ts–GFP[24] and native peroxisomal protein Pex22, were overexpressed and remained soluble (Extended Data Fig. 8a–c). They did not significantly affect proteasome activity (Extended Data Fig. 8d–g). To better mimic the stress caused by delayed mitochondrial import, we induced Mix17 in the presence of CCCP. The stimulatory effects of import inhibition and Mix17 overexpression on proteasome activity were additive (Fig. 3c and Extended Data Fig. 8h). Thus, proteins that were delayed in their transport to mitochondria—because of overexpression or the absence of mitochondrial targeting signals or chemical import inhibition—stimulated proteasomal activity.

No proteasomal activity stimulation by Mix17 was observed in cells lacking Irc25 or Poc4 (Fig. 3d and Extended Data Fig. 9a, b). Indeed, proteasomal subunits were not upregulated (in contrast to Irc25 and Poc4; Fig. 1f and Supplementary Table 3) and the purified proteasome was not stimulated *in vitro* by Pet191$_{Flag}$ or Mix17$_{Flag}$ (Extended
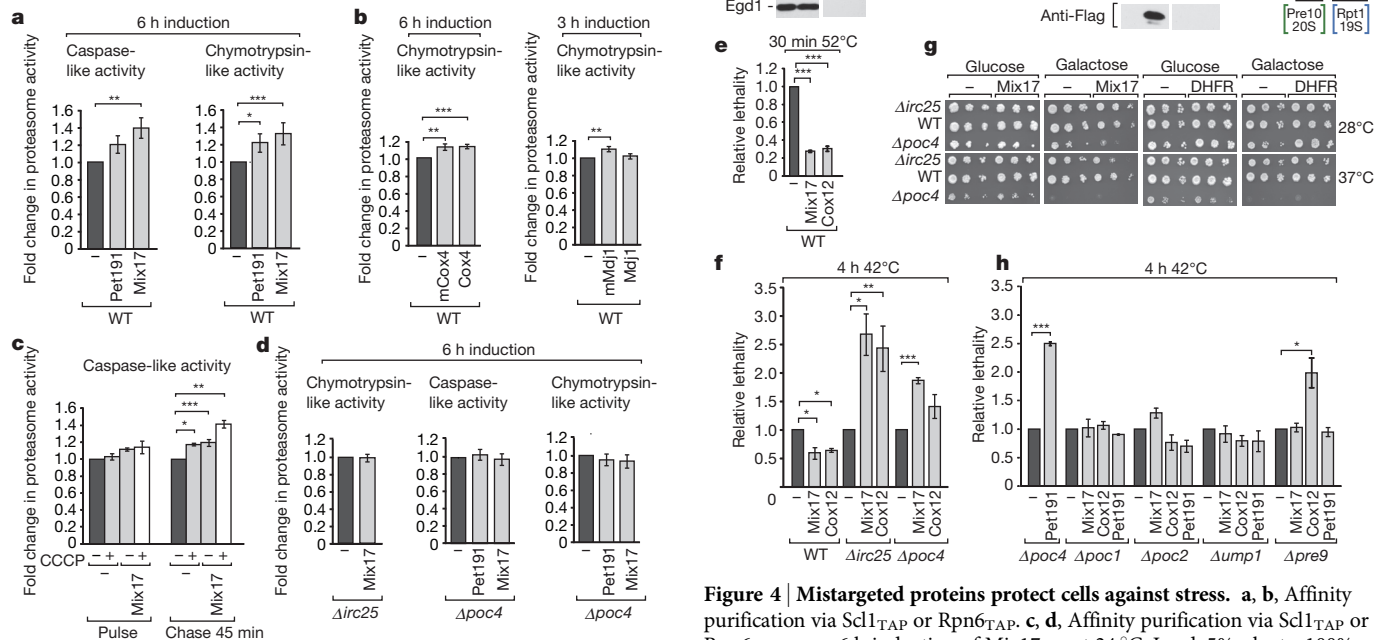

Figure 3 | Mitochondrial precursor proteins stimulate proteasomal activity. **a,** Proteasomal activity upon expression of Pet191$_{Flag}$ or Mix17$_{Flag}$ at 24 °C. Mean ± s.e.m., $n = 6$. **b,** Proteasomal activity upon expression of mCox4$_{Flag}$ or Cox4$_{Flag}$ at 28 °C and mMdj1$_{Flag}$ or Mdj1$_{Flag}$ at 24 °C. Mean ± s.e.m., $n = 5$. **c,** Proteasomal activity upon Mix17$_{Flag}$ expression and CCCP treatment. Mean ± s.e.m., $n = 3$. **d,** Proteasomal activity in cells lacking Irc25 or Poc4 upon expression of Pet191$_{Flag}$ or Mix17$_{Flag}$ at 24 °C. Mean ± s.e.m., $n = 3$ (caspase-like activity), $n = 4$ (chymotrypsin-like activity). **a–c,** *$P < 0.05$; **$P < 0.03$; ***$P < 0.01$. WT, wild-type.

Figure 4 | Mistargeted proteins protect cells against stress. **a, b,** Affinity purification via Scl1$_{TAP}$ or Rpn6$_{TAP}$. **c, d,** Affinity purification via Scl1$_{TAP}$ or Rpn6$_{TAP}$ upon 6 h induction of Mix17$_{Flag}$ at 24 °C. Load, 5%; eluate, 100%. The empty vector eluate was set to 1. Mean ± s.e.m, $n = 4$ (**c**), $n = 5$ (**d**). **e, f,** Lethality upon overexpression of mitochondrial proteins. Mean ± s.e.m, $n = 4$ (**e**), $n = 3$ (**f**). Mean lethality values of cells with empty vector: wild type, 19% (**e**) or wild type, 9.54%; *Δirc25*, 9.01%; *Δpoc4*, 7.13% (**f**), were set to 1. **g,** Cells were subjected to consecutive tenfold dilutions. **h,** Lethality of cells. Mean ± s.e.m, $n = 3$. Mean lethality values of cells with empty plasmid: *Δpoc4*, 5.33%; *Δpoc1*, 11.02%; *Δpoc2*, 11.68%; *Δump1*, 9.91%; *Δpre9*, 10.74%, were set to 1. **c–e, f, h,** *$P < 0.05$; **$P < 0.03$; ***$P < 0.01$. WT, wild-type. **a–d,** Uncropped blots are in Supplementary Information Fig. 1.

Data Fig. 9c). The 26S proteasome is assembled from components, including the 20S catalytic core and the 19S regulatory particle[10–12]. Scl1$_{TAP}$ (core subunit) and Rpn6$_{TAP}$ (regulatory subunit) were used to purify the 20S core (represented by Pre10) and the 26S proteasome that comprised the core and regulatory particles (represented by Rpt1 and Rpt5) (Fig. 4a, b and Extended Data Fig. 9d). The proteasomal subunits were eluted more efficiently with Scl1$_{TAP}$ and Rpn6$_{TAP}$ upon Mix17 overexpression (Fig. 4c, d). Thus, the stimulatory effect was associated with proteasomal assembly.

Interestingly, heat-shock conditions revealed a gain of stress resistance upon mitochondrial precursor appearance (Fig. 4e, f and Extended Data Fig. 10a). In the cells without Irc25 or Poc4, Mix17$_{Flag}$ and Cox12$_{Flag}$ exerted toxicity under temperature-stress conditions (Fig. 4f). The induction of Pet191$_{Flag}$, Mix17$_{Flag}$ and Cox12$_{Flag}$ but not DHFR$_{Flag}$ in these cells led to a synthetic growth defect (Fig. 4g and Extended Data Fig. 10b). Cells without Poc1, Poc2, Ump1 or the proteasomal component Pre9 did not exhibit an increase in lethality (Fig. 4h). Thus, the response that was activated by protein mistargeting relied on proteasome assembly governed by the Irc25–Poc4 complex.

The mitochondrial protein import slowdown and precursor over-accumulation stress triggers processes that maintain cellular proteostasis. This includes the inhibition of protein synthesis and activation of the proteasome that originates from mistargeted proteins (Extended Data Fig. 10c). This unfolded protein response activated by protein mistargeting (UPRam) exerts a protective effect on the cells. UPRam differs from UPRmt, another response pathway to compromised mitochondrial function present in higher Eukaryotes, because UPRmt acts by regulating the abundance of chaperones and proteases inside mitochondria[25–27]. Proteasomal activation and mild mitochondrial dysfunction have both been strongly implicated in the positive regulation of longevity[28–30]. The newly identified UPRam causally links defects in mitochondrial biogenesis with proteasome activity, thus providing a possible explanation for the beneficial effects of mild mitochondrial dysfunction on ageing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. & Pfanner, N. Importing mitochondrial proteins: machineries and mechanisms. *Cell* **138**, 628–644 (2009).
2. Endo, T., Yamano, K. & Kawano, S. Structural insight into the mitochondrial protein import system. *Biochim. Biophys. Acta* **1808**, 955–970 (2011).
3. Neupert, W. & Herrmann, J. M. Translocation of proteins into mitochondria. *Annu. Rev. Biochem.* **76**, 723–749 (2007).
4. van der Laan, M., Hutu, D. P. & Rehling, P. On the mechanism of preprotein import by the mitochondrial presequence translocase. *Biochim. Biophys. Acta* **1803**, 732–739 (2010).
5. Chacinska, A. *et al.* Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins. *EMBO J.* **23**, 3735–3746 (2004).
6. Stojanovski, D. *et al.* Mitochondrial protein import: precursor oxidation in a ternary complex with disulfide carrier and sulfhydryl oxidase. *J. Cell Biol.* **183**, 195–202 (2008).
7. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
8. Bragoszewski, P., Gornicka, A., Sztolsztener, M. E. & Chacinska, A. The ubiquitin-proteasome system regulates mitochondrial intermembrane space proteins. *Mol. Cell. Biol.* **33**, 2136–2148 (2013).
9. Brodsky, J. L. Cleaning up: ER-associated degradation to the rescue. *Cell* **151**, 1163–1167 (2012).
10. Ciechanover, A. Intracellular protein degradation: from a vague idea through the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Biochim. Biophys. Acta* **1824**, 3–13 (2012).
11. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
12. Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).
13. Bragoszewski, P. *et al.* Retro-translocation of mitochondrial intermembrane space proteins. *Proc. Natl Acad. Sci. USA* **112**, 7713–7718 (2015).
14. Kusmierczyk, A. R., Kunjappu, M. J., Funakoshi, M. & Hochstrasser, M. A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nature Struct. Mol. Biol.* **15**, 237–244 (2008).
15. Le Tallec, B. *et al.* 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals. *Mol. Cell* **27**, 660–674 (2007).
16. Hirano, Y. *et al.* Cooperation of multiple chaperones required for the assembly of mammalian 20S proteasomes. *Mol. Cell* **24**, 977–984 (2006).
17. Ramos, P. C., Hockendorff, J., Johnson, E. S., Varshavsky, A. & Dohmen, R. J. Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. *Cell* **92**, 489–499 (1998).
18. Hanssum, A. *et al.* An inducible chaperone adapts proteasome assembly to stress. *Mol. Cell* **55**, 566–577 (2014).
19. Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770–774 (2000).
20. Fischer, M. *et al.* Protein import and oxidative folding in the mitochondrial intermembrane space of intact mammalian cells. *Mol. Biol. Cell* **24**, 2160–2170 (2013).
21. Frazier, A. E. *et al.* Pam16 has an essential role in the mitochondrial protein import motor. *Nature Struct. Mol. Biol.* **11**, 226–233 (2004).
22. Chacinska, A. *et al.* Mitochondrial presequence translocase: switching between TOM tethering and motor recruitment involves Tim21 and Tim17. *Cell* **120**, 817–829 (2005).
23. Bennett, E. J., Bence, N. F., Jayakumar, R. & Kopito, R. R. Global impairment of the ubiquitin–proteasome system by nuclear or cytoplasmic protein aggregates precedes inclusion body formation. *Mol. Cell* **17**, 351–365 (2005).
24. Escusa-Toret, S., Vonk, W. I. & Frydman, J. Spatial sequestration of misfolded proteins by a dynamic chaperone pathway enhances cellular fitness during stress. *Nature Cell Biol.* **15**, 1231–1243 (2013).
25. Haynes, C. M. & Ron, D. The mitochondrial UPR — protecting organelle protein homeostasis. *J. Cell Sci.* **123**, 3849–3855 (2010).
26. Zhao, Q. *et al.* A mitochondrial specific stress response in mammalian cells. *EMBO J.* **21**, 4411–4419 (2002).
27. Nargund, A. M., Pellegrino, M. W., Fiorese, C. J., Baker, B. M. & Haynes, C. M. Mitochondrial import efficiency of ATFS-1 regulates mitochondrial UPR activation. *Science* **337**, 587–590 (2012).
28. Vilchez, D. *et al.* RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. *Nature* **489**, 263–268 (2012).
29. Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144**, 79–91 (2011).
30. Houtkooper, R. H. *et al.* Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* **497**, 451–457 (2013).

Supplementary Information is available in the online version of the paper.

Author Contributions P.B. and S.W. are joint second authors. L.W., U.T., P.B., M.E.S., A.V., P.C., S.M. and E.J. performed and analysed biochemical experiments. P.B. and M.L. performed RNA-seq and analyses. S.W. and S.O. performed the mass spectrometric measurements and analyses. A.C., B.W., M.K. and A.D. analysed and supervised the study. A.C and B.W. conceived the project. All authors interpreted the experiments. A.C. wrote the manuscript with the input of other authors.

Author Information RNA-seq data have been submitted to the ArrayExpress database under accession number E-MTAB-3588. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD001495. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (achacinska@iimcb.gov.pl) or B.W. (bettina.warscheid@biologie.uni-freiburg.de).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains. Yeast *Saccharomyces cerevisiae* strains are derivatives of YPH499 (*MATa, ade2-101, his3-Δ200, leu2-Δ1, ura3-52, trp1-Δ63, lys2-801*) or BY4741 (*MATa, his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*). The temperature-sensitive *mia40-4int* (Fomp2-7int; 305), *mia40-4* (YPH-fomp2-7; 176), *mia40-3* (YPH-BG-fomp2-8; 178), *mia40-F311E* (660) and corresponding wild-type (398) strains were described previously[5,6,8]. The *mia40-4int* was generated by integration of the plasmid-borne *mia40-4* allele into the genome[6]. This mutant showed a stronger decrease in the accumulation of intermembrane space proteins than *mia40-4* (refs 5, 6, 8), but due to a single genomic copy of *mia40* was more suitable for SILAC. The SILAC strains are *arg8::kanMX4* derivatives of YPH499 (524) and *mia40-4int* (305) and are referred as to WT[S] (654) and *mia40-4int*[S] (558; a gift from Bernard Guiard). The *pam16-1* (YPH-BG-mia1-1; 733), *pam16-3* (YPH-BG-mia1-3; 734), *pam18-1* (YPH-BG-Mdj3-66; 739), *tim17-4* (YPH-BG17-9d; 83), *tim17-5* (YPH-BG17-21-7; 85) and corresponding wild-type (738, 524) strains were described previously[21,22,31]. Strains expressing *PRE3* (843) or *RPN6* (905) with a TAP tag and deletion strains of *PRE9*, *POC1*, *POC2*, *IRC25*, *POC4*, *UMP1* in the BY4741 genetic background were purchased from Euroscarf. The strain expressing *SCL1* with TAP tag was a gift from Michael Glickman (916). To rescue *mia40-4int*, the wild-type and *mia40-4int* yeast cells were transformed with a plasmid containing *MIA40* under an endogenous promoter (pGB8220; 85p). To overproduce Mia40, yeast cells were transformed with the TRP1-containing plasmid (pAC8-3, 97p) harbouring *MIA40* (YKL195W) under control of the *GAL10* promoter. The plasmid for Ubc9ts–GFP overproduction was purchased from Addgene (314p)[24]. To express *PET191* (YJR034W), *MIX17* (YMR002W), *COX12* (YLR038C), *POC1* (YLR199C), *POC2* (YKL206C), *IRC25* (YLR021W), *POC4* (YGL187C), *COX4* (YGL187C), *MDJ1* (YFL016C), *PEX22* (YAL055W) and DHFR encoding mouse dihydrofolate reductase or its destabilized version DHFRds in frame with the FLAG tag under control of the *GAL10* promoter, yeast cells were transformed with URA3- or LEU2-containing pAG1 (53p), pAG2 (54p), pEJ1 (334p), pAG3 (55p), pPCh6 (320p), pPCh1 (315p), pMS126 (151p), pMS127 (152p), pMS120 (145p), pMS121 (146p), pUT9 (333p), pMS128 (162p), pMS119 (144p) plasmids. Mix17 (YMR002W) was previously known as Mic17 (refs 8, 32, 33). The mMdj1 and mCox4 constructs correspond to 17–511 and 25–155 amino acid residues of Mdj1 and Cox4, respectively. The sequence for DHFRds contains three destabilizing amino acid substitutions C7S, S42C, N49C (gift from Wolfgang Voos[34,35]).

Growth conditions. The strains were grown on minimal synthetic medium (0.67% (w/v) yeast nitrogen base, 0.079% (w/v) CSM amino acid mix) containing respiratory carbon source (3% (v/v) glycerol supplemented with 0.05–0.2% (v/v) glucose) or fermentable carbon source (2% (v/v) sucrose, 2% (v/v) galactose, 2% (v/v) glucose). To induce the *GAL10* promoter for the expression of *MIA40,* cells were grown on minimal selective medium with 1.5% (v/v) glycerol supplemented with 0.5% (v/v) galactose. To induce the *GAL10* promoter for the expression of other proteins, cells were grown on minimal selective medium with 2% (v/v) galactose. In the experiments in which Mia40 and Pet191 were simultaneously overproduced, cells were grown on minimal selective medium with 2% (v/v) galactose. The yeast strains were usually grown at permissive temperature (19 °C or 24 °C) to the early logarithmic growth phase and analysed before or after a shift to a restrictive temperature (37 °C).

RNA-seq. Wild-type and *mia40-4int* cells were grown at 19 °C in YPG (1% (w/v) yeast extract, 2% (w/v) peptone, 3% (v/v) glycerol) medium to the mid-logarithmic phase (D_{600nm} (OD600) of ~0.65) in three biological replicates per genotype. RNA was extracted with MagNALyser/MagNA Pure Compact RNA Isolation Kit (Roche) from frozen yeast preserved with RNAlater (Ambion). DNA was digested with TURBO DNA-free Kit (Life Technologies). The RNA quality and concentration were tested with Agilent 2100 Bioanalyzer and Qubit RNA BR Assay (Life Technologies). RNA was spiked with ERCC RNA Spike-In Mix (Life Technologies) and rRNA depleted with Ribo-Zero Gold rRNA Removal Kit (Illumina). RNA-seq libraries were prepared with Ion Total RNA-Seq Kit v2 (Life technologies) with 2 + 10 cycles amplification. Equimolar pool of barcoded libraries was used to prepare sequencing template with Ion PI Template OT2 200 Kit v3 (Life Technologies), followed by sequencing with Ion Proton on Ion PI v2 chip (Life Technologies). The reads were adaptor trimmed and aligned using Torrent Suite and Partek Flow as recommended by Life Technologies. The reads were quality trimmed by Torrent Suite and resulting files were aligned to the Ensembl *S. cerevisiae* R64 reference genome (GCA_000146045.2) by TopHat2 (ref. 36). Not aligned reads were submitted to the Bowtie2 (ref. 37) alignment with the 'very sensitive' setting. Alignments were combined and coordinate sorted with SAMtools[38]. Differential expression analysis was performed with Partek Genomics

Suite (Partek). Transcripts were quantified to the Ensembl R64 GTF annotation file and resulting RPKM values were used for statistical analysis in Partek ANOVA (General Linear Model).

SILAC labelling and LC/MS sample preparation. Yeast strains of WT[S] and *mia40-4int*[S] were grown at 19 °C on respiratory minimal liquid medium (0.67% (w/v) yeast nitrogen base, 0.068% (w/v) CSM amino acid mix lacking arginine and lysine, 3% (v/v) glycerol, 0.02% (w/v) L-proline) supplemented with 18.6 mg l^{-1} [$^{13}C_6/^{15}N_4$] L-arginine and 19.35 mg l^{-1} [$^{13}C_6/^{15}N_2$] L-lysine (Sigma-Aldrich) or non-labelled L-arginine and L-lysine. Three separate cultures were grown and differentially labelled using SILAC amino acids including a label-switch. Following mixing of differentially SILAC-labelled WT[S] and *mia40-4int*[S] cells and cell lysis, proteins were acetone-precipitated and analysed using a gel-free and a gel-based approach. For the gel-free approach, proteins were redissolved in 60% MeOH/20 mM NH_4HCO_3 (pH 7.8) followed by reduction and alkylation of cysteines and tryptic digest as described[39]. For the gel-based approach, proteins were resuspended in urea buffer (7 M urea/2 M thiourea/30 mM Tris-HCl, pH 8.5) and separated (15 μg protein per replicate) on a 4–12% NuPAGE Bis-Tris gel. Lanes were cut into 15 slices, destained and washed as reported[40] followed by reduction and alkylation of cysteines[39]. Proteins were in-gel digested with trypsin and peptides reconstituted in 0.1% TFA.

LC/MS analysis. Peptides obtained by in-solution digestion were analysed in triplicates per biological replicate using an RSLCnano/LTQ-Orbitrap XL system (Thermo Scientific, Bremen, Germany) as described[39] with the modification that peptides were separated with a 150-min gradient (4–35% acetonitrile in 0.1% formic acid). Peptides derived from the gel were analysed on an RSLCnano/Orbitrap Elite system using a 45-min gradient and parameters as follows: survey scan range, *m/z* 370–1,700; resolution of 120,000 (*m/z* 400); top15 method (CID); automatic gain control and maximum fill time, 1×10^6 and 200 ms for the orbitrap, 5×10^3 and 150 ms for the linear ion trap; normalized collision energy, 35%; activation *q*, 0.25; activation time, 10 ms; dynamic exclusion, 45 s.

MS data analysis. Data were processed using MaxQuant/Andromeda (version 1.3.0.5)[41,42] and searched against the *Saccharomyces* Genome Database (http://www.yeastgenome.org). LC/MS data from mitochondrial samples were included to support peptide identifications. The database search was performed with default settings including deamination of asparagine and Pro6 as variable modifications. Protein identifications were based on at least one unique peptide. Mass tolerances for precursor and fragment ions were 6 p.p.m. and 0.5 Da, respectively. A false discovery rate of 1% was applied on peptide and protein level. SILAC ratios were calculated based on unique peptides using MaxQuant default settings and "match between runs". SILAC ratios for proteins were log-transformed and mean log_{10} ratios across all replicates and *P* values (two-sided *t*-test) were determined. Protein groups exhibiting a posterior error probability (PEP) < 0.01, a *P* value <0.05 and a minimum fold change ±1.5 were classified as up- or downregulated in *mia40-4int*[S] cells. Variances of the mean log_{10} SILAC ratios were consistent across all three biological replicates yielding uniform data. Sample size, data analysis approach and the statistical test chosen fulfil the requirements for quantitative proteome analyses employing similar LC/MS instrumentation and methods. Statistical methods to determine sample size were not applied. The acquired data were normally distributed, thus meeting the assumption of the statistical test. All raw data and original MaxQuant result files have been deposited to the ProteomeXchange Consortium[43] via the PRIDE partner repository[44] with the data set identifier PXD001495. GO term enrichment analysis of proteins classified as upregulated or downregulated was performed using the BiNGO[45] plug-in 2.44 for Cytoscape 2.8 (ref. 46) and the yeast genome as background. Using the built-in Benjamini–Hochberg procedure, raw *P* values calculated for GO enrichment analysis were corrected for multiple testing. GO terms with a corrected *P* value <0.05 were considered enriched.

Proteasome activity assays. Proteasome activity assays were carried out as previously described[47] with few modifications. 20 $D_{600\,nm}$ (OD600) units of cells were shaken in lysis buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10% glycerol, 2 mM ATP, 2 mM PMSF, 1 mM DTT) with glass beads (Sigma-Aldrich) for 10–20 min at 4 °C. After a clarifying spin (20,000*g*, 15 min, 4 °C), protein concentration was determined by Bradford assay with bovine serum albumin as standard. Activity assays were performed in a final volume of 200 μl of lysis buffer with 50 μg of soluble total protein extracts in a 96-well plate by adding 100 μM Suc–Leu–Leu–Val–Tyr–AMC peptide substrate (chymotrypsin-like activity; Bachem, I-1395) or 100 μM Ac–NIe–Pro–NIe–Asp–AMC (caspase-like activity; Bachem, I-1850). Fluorescence (excitation wavelength 380 nm, emission wavelength 460 nm) was measured every 5 min for 1–2 h at 25 °C using a microplate fluorometer (Infinite M1000, Tecan). For *in vivo* inhibition of proteasome activity, cells were grown as described previously[8]. Cultures were supplemented with either 75 μM MG132 or DMSO.

Stress assays. Cells were grown to the early logarithmic phase and expression of proteins was induced for 6 h at 24 °C in minimal medium (0.67% yeast nitrogen base, 0.077% amino acid supplement mixture) containing 2% (v/v) galactose. Only cultures with similar D_{600nm} (OD600) were compared for their viability after heat shock. One D_{600nm} (OD600) unit of yeast cells was washed with PBS (137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4). Cells were stained with 3 µg ml^{-1} propidium iodide in PBS for 15 min at room temperature in dark and analysed within 1 h by flow cytometry (FACSCalibur). 10,000 cells for each sample were measured.

Mitochondria procedures and cell fractionation. Yeast were usually grown on YPG medium. The yeast with overproduced Mia40 were grown on YPGal (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) galactose). Mitochondria were isolated by differential centrifugation according to the standard method[48]. Mitochondria were solubilized in Laemmli buffer with 50 mM DTT (reducing) or 50 mM iodoacetamide (non-reducing conditions), and analysed by SDS–PAGE followed by western blotting. [^{35}S]methionine radiolabelled precursors were prepared by in vitro transcription (mMESSAGE mMACHINE, Ambion) and translation (rabbit reticulocyte lysate system, Promega) and imported into the isolated mitochondria as previously described[6]. Import experiments were performed in at least two different mitochondrial isolations (biological replicates) and repeated at least three times (technical replicates). For fractionation, yeast were grown on minimal medium with fermentable carbon source (2% glucose) to D_{600nm} (OD600) 0.5, induced for 6 h or overnight by 2% galactose for overexpression of *MIA40* and 35 D_{600nm} (OD600) units were fractionated by differential centrifugation[8].

Mitochondrial precursor protein analysis. Yeast cells were grown in YPGal with addition of 0.2% glucose at 19 °C. Upon consumption of glucose, yeast were grown for additional 2–3 h to D_{600nm} (OD600) 1.5–3. Cultures were then supplemented with either 40 µM CCCP or DMSO. After 1.5 h, samples were analysed by SDS–PAGE and western blotting. For pulse-chase experiments, upon consumption of glucose, yeast were grown for an additional 1 h to D_{600nm} (OD600) 1.1–1.3. Cultures were then supplemented with either 10 µM CCCP or DMSO for the pulse. For pulse-chase experiments with CCCP in combination with expression of Mix17$_{Flag}$ cells were grown at 24 °C in minimal selective medium containing 2% (v/v) sucrose to logarithmic phase. Cells were washed and resuspended in YPGal medium and grown for 4 h at 24 °C. Cells were shifted for 15 min at 19 °C before supplementing the culture with either 10 µM CCCP or DMSO and samples were chased at 19 °C.

Aggregation assay. Cells were grown for 4 h at 28 °C with either 75 µM MG132 or the corresponding volume of DMSO[8]. Then 10 D_{600nm} (OD600) units of cells were shaken in lysis buffer (30 mM Tris-HCl pH 7.4, 20 mM KCl, 5 mM EDTA, 0.5 mM PMSF, 0.5% Trition X-100) with glass beads for 10 min at 4 °C. After clarifying centrifugation (4,000g for 10 min, 4 °C), total (T) samples were separated into pellet (P) of aggregated proteins and supernatant (S) by ultracentrifugation for 1 h at 125,000g. Total and supernatant fractions were precipitated by 10% trichloroacetic acid.

Affinity purifications. For purification of Flag-tagged proteins, cells were grown at 28 °C in 3% (v/v) glycerol minimal medium with 0.5% (v/v) galactose. Cells (100U D_{600nm} (OD600)) were shaken in buffer A (20 mM Tris-HCl pH 7.4, 300 mM NaCl, 2 mM PMSF) with glass beads for 20 min at 4 °C. The solution was supplemented with 1% (w/v) digitonin and incubated for 15 min on ice. After clarifying centrifugation (20,000g, 15 min, 4 °C), supernatants were incubated with anti-Flag M2 affinity gel (Sigma-Aldrich) at 4 °C for 1 h. The column was washed with buffer A, followed by elution with 0.1 M glycine pH 3.5 for 5 min at room temperature. The pH of the elution fraction was adjusted to pH 7.4.

For proteasome complex purification, 100 D_{600nm} (OD600) units of cells were shaken in buffer B (50 mM Tris-HCl pH 7.4, 10 mM MgCl$_2$, 10% glycerol, 2 mM ATP, 2 mM PMSF, 1 mM DTT) with glass beads for 10 min at 4 °C. After clarifying centrifugation (20,000g, 15 min, 4 °C) supernatants were incubated with immunoglobulin G Sepharose (GE Healthcare) for 1 h at 4 °C. The column was washed and resuspended in 100 µl of buffer B (for purified proteasome activity assay) or proteins were eluted by Laemmli buffer with 50 mM DTT. Proteasome components co-purified with TAP-tagged proteins were quantified and the ratio of signals from cells with Mix17 versus empty plasmid was calculated and normalized to the efficiency of Scl1$_{TAP}$ or Rpn6$_{TAP}$ recovery in the eluate.

Ribosome profile analysis. Yeast were grown on YPGal to early logarithmic phase and shifted to 37 °C. Cells were disrupted in lysis buffer (40 mM Tris-HCl pH 7.4, 100 mM NaCl, 30 mM MgCl$_2$, 100 µg ml^{-1} cycloheximide, 300 µg ml^{-1} heparin, 1 mM DTT, 0.5 mM PMSF, 2 µM pepstatin A, 2 µg ml^{-1} chymostatin, 1 µM leupeptin, 2 mM benzamidie HCl and 5 µM chymostatin) by shaking with glass beads followed by Triton X-100 addition to 0.2% and incubation for 15 min on ice. After clarifying spin (20,000g, 10 min, 4 °C), samples were separated on 10–50% (w/v) sucrose gradients in polysome buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 5 mM MgCl$_2$, 100 µg ml^{-1} cycloheximide, 2 mM DTT; 222,000g,

2 h, 4 °C). Gradients were fractionated with a continuous absorbance measurement at 254 nm. The peak areas (volume (ml) × absorbance (mAU)) were quantified with Unicorn 5.2 (GE Healthcare).

Northern blot and qPCR assays. Total RNA was extracted from cells as described[49]. For northern blot 5 µg of RNA was separated on agarose or acrylamide denaturing gels, transferred to Hybond N+ (GE Healthcare) membrane and immobilized by the UVP CL-1000 crosslinker. Oligonucleotides used for RNA hybridizations (25S rRNA (007) 5′-CTCCGCTTATTGATATGC; 18S rRNA (008) 5′-CATGGCTTAATCTTTGAGAC; 5.8S rRNA (017) 5′-GCGTTGTTC ATCGATGC; 5S rRNA 5′-CTACTCGGTCAGGCTC) were labelled with [γ-^{32}P]ATP using T4 polynucleotide kinase (New England Biolabs). All northern blot steps were performed using PerfectHyb Plus Hybridization Buffer (Sigma-Aldrich). Radioactive signals were detected by digital autoradiography. For qPCR analysis RNA extracts were treated with TURBO DNase (Ambion) and re-isolated. 1 µg of RNA was reverse transcribed using SuperScript III First-Strand Synthesis SuperMix (Invitrogen) with 2.5 µM oligo(dT)$_{20}$ and 1.8 µM random hexamers as primers. Quantitative PCR reactions were performed using 7900HT Real-Time PCR (LifeTechnologies) and SensiFAST SYBR Hi-ROX Kit (Bioline) with standard protocol in triplicates. Standard curves were generated by the serial dilution of the pooled cDNA. To normalize nonspecific variations, the normalization factor was calculated as the geometric mean of transcript levels of 3 control genes, *ALG9*, *FBA1* and *TUB2* that were selected on the basis of stability in transcriptomic and proteomic data sets and published observations[50–52]. Primers used for qPCR: *ALG9* 5′-CACGGATAGTGGCTTTGGTGAACAAT TAC, 5′-TATGATTATCTGGCAGCAGGAAAGAACTTGGG; *FBA1* 5′-ACGA AGGTCAAAATGCTTCCAT, 5′-TGGCACAGTGGTCAGAGTGTAAG; *IRC25* 5′-CACTGGCAGGAAACCTAGGAAA, 5′-CTCCCAACTTGGTAACCACTT GA; *POC4* 5′-CTCTACCTGCAGACGATCGTTCT, 5′-GGGAATGGCGTAGT AATAACATGA; *TUB2* 5′-ATCTTGTCCCATTCCCACGTT, 5′-TTGCTGTGT TAATTCAGGGACAGT.

Statistical analysis. For statistical analysis, a two-tailed, paired t-test was used. For statistical analysis of SILAC, polysome profiles, northern blots and qPCR, a two-tailed, unpaired t-test, assuming unequal variance, was performed. A P value < 0.05 was considered significant. Proteasomal activity was calculated as a slope value and data are represented in a fold change compared to wild-type strain or wild-type containing the empty plasmid (with the exception for the pulse-chase experiments, in which the CCCP-treated samples were compared to the DMSO-treated control). Represented fold changes were the mean of fold changes obtained from independent biological replicates ± standard error of the mean (s.e.m.). ImageQuant was used for the quantification of western blotting results. The sample size used was estimated based on previous experiments and statistical tests were chosen based on published data with comparable methodology. Experiments have been replicated at least three times.

Other details. Volcano plot distribution of transcripts and proteins was based on the GO Slim terms provided by the *Saccharomyces* genome database (SGD). Protein extracts were prepared from amount of cells corresponding to 0.2–0.4 D_{600nm} (OD600) units by alkaline lysis[53] and resuspended in Laemmli buffer containing 50 mM DTT. To assess protein synthesis, cells were grown in Met-free minimal medium before incubating with [^{35}S]-labelled methionine and cysteine (EXPRESS; Perkin Elmer) that were added to a final concentration of 6 µCi ml^{-1} for 20 min to 2 h. The samples were analysed by SDS–PAGE and digital autoradiography. Proteins were separated by SDS–PAGE on 15%, 12% or 10% gels. Primary antibodies were custom-raised in rabbits and individually controlled for specificity. Commercially available antibodies for mouse monoclonal anti-ubiquitin (SC-8017, Santa Cruz), anti-Flag M2 (F1804, Sigma-Aldrich), rabbit polyclonal anti-Rpt5 (PW8245, Enzo Life Science), anti-Rpt1 (PW8255, Enzo Life Science), anti-alpha7/Pre10 (PW8110, Enzo Life Science) and anti-Hsp104 (ADI SPA 1040, Enzo Life Science) were used. Chemiluminescence protein signals were detected by use of X-ray films. Autoradiography signals were processed with ImageQuant (GE Healthcare). The images were processed digitally using Adobe Photoshop CS4. In some figures, non-relevant gel parts were excised digitally. Protein concentration was determined using Roti-Quant (Carl Roth GmbH) with bovine serum albumin as a standard. The nomenclature of proteins is according to the SGD.

31. Truscott, K. N. et al. A J-protein is an essential subunit of the presequence translocase-associated protein import motor of mitochondria. *J. Cell Biol.* **163,** 707–713 (2003).
32. Böttinger, L. et al. In vivo evidence for cooperation of Mia40 and Erv1 in the oxidation of mitochondrial proteins. *Mol. Biol. Cell* **23,** 3957–3969 (2012).
33. Pfanner, N. et al. Uniform nomenclature for the mitochondrial contact site and cristae organizing system. *J. Cell Biol.* **204,** 1083–1086 (2014).
34. Vestweber, D. & Schatz, G. Point mutations destabilizing a precursor protein enhance its post-translational import into mitochondria. *EMBO J.* **7,** 1147–1151 (1988).

35. Schreiner, B. et al. Role of the AAA protease Yme1 in folding of proteins in the intermembrane space of mitochondria. Mol. Biol. Cell 23, 4335–4346 (2012).

36. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).

37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359 (2012).

38. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

39. Lytovchenko, O. et al. The INA complex facilitates assembly of the peripheral stalk of the mitochondrial F_1F_o-ATP synthase. EMBO J. 33, 1624–1638 (2014).

40. Wiese, S. et al. Proteomics characterization of mouse kidney peroxisomes by tandem mass spectrometry and protein correlation profiling. Mol. Cell. Proteomics 6, 2045–2057 (2007).

41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnol. 26, 1367–1372 (2008).

42. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10, 1794–1805 (2011).

43. Vizcaíno, J. A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nature Biotechnol. 32, 223–226 (2014).

44. Vizcaíno, J. A. et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 41, D1063–D1069 (2013).

45. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448–3449 (2005).

46. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).

47. Kruegel, U. et al. Elevated proteasome capacity extends replicative lifespan in Saccharomyces cerevisiae. PLoS Genet. 7, e1002253 (2011).

48. Meisinger, C., Pfanner, N. & Truscott, K. N. Isolation of yeast mitochondria. Methods Mol. Biol. 313, 33–39 (2006).

49. Tollervey, D. & Mattaj, I. W. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. EMBO J. 6, 469–476 (1987).

50. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT–PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, research0034 (2002).

51. Teste, M. A., Duquenne, M., Francois, J. M. & Parrou, J. L. Validation of reference genes for quantitative expression analysis by real-time RT–PCR in Saccharomyces cerevisiae. BMC Mol. Biol. 10, 99 (2009).

52. Cankorur-Cetinkaya, A. et al. A novel strategy for selection and validation of reference genes in dynamic multidimensional experimental design in yeast. PLoS ONE 7, e38351 (2012).

53. Kushnirov, V. V. Rapid and reliable protein extraction from yeast. Yeast 16, 857–860 (2000).
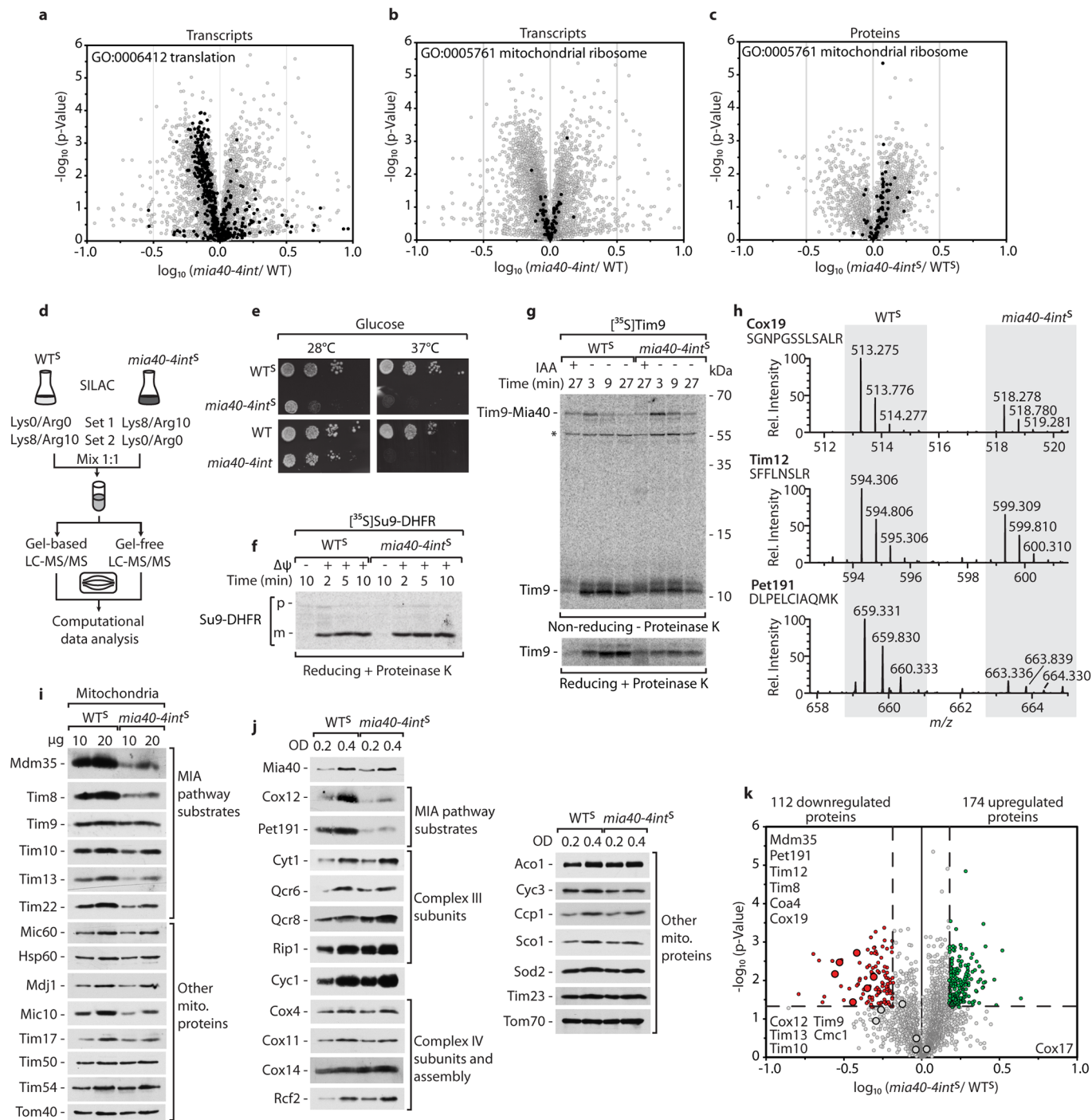
Extended Data Figure 1 | Transcriptome and proteome analysis of WT and *mia40-4int* cells. a, Distribution of transcripts quantified by RNA-seq in wild-type and *mia40-4int* cells based on the GO term for translation provided by the *Saccharomyces* genome database. **b, c**, Distribution of transcripts and proteins based on the GO term for mitochondrial ribosome provided by the *Saccharomyces* genome database. **d**, SILAC-based strategy for the quantitative analysis of alterations in the proteome of yeast cells bearing the *mia40-4int* mutation. Derivatives of *mia40-4int* and wild type (*mia40-4int*[S] and WT[S]) were used for SILAC. The *mia40-4int*[S] cells grown at permissive temperature 19 °C in heavy medium containing Lys8 and Arg10 were mixed in equal ratio with WT[S] cells grown in the light medium containing Lys0 and Arg0 for set 1 (two biological replicates) or vice versa for set 2 (one additional biological replicate). **e**, The *mia40-4int*[S] strain and the parental *mia40-4int* strain exhibited a similar temperature-sensitive phenotype. Both mutants and corresponding wild-type strains were subjected to consecutive tenfold dilutions, spotted on YPD plates and grown at the indicated temperatures. **f**, The *mia40-4int*[S] strain showed no defect in the import of matrix-targeted Su9-DHFR as expected for *mia40-4int*. The [35S]-labelled precursor of Su9-DHFR was incubated with isolated mitochondria from *mia40-4int*[S] and WT[S] strains for the indicated time points. **g**, Import of the MIA substrate, Tim9, was decreased in *mia40-4int*[S] as expected for *mia40-4int*. [35S]-labelled

precursor of Tim9 was incubated with the isolated mitochondria from *mia40-4int*[S] and WT[S] strains for the indicated time points. **f, g**, An excess of non-imported precursor was removed by the treatment of mitochondria with proteinase K. Samples were analysed by reducing and non-reducing SDS–PAGE followed by autoradiography. **h**, Sections of MS survey spectra of SILAC-encoded peptides for the MIA pathway substrates Cox19, Tim12 and Pet191 exhibiting decreased levels in *mia40-4int*[S] cells. MS survey spectra were acquired from the experimental set 1 as depicted in Extended Data Fig. 1d. **i**, Mitochondria were isolated from *mia40-4int*[S] and corresponding WT[S] grown at 19 °C and analysed by western blotting. **j**, WT[S] and *mia40-4int*[S] yeast were grown to the stationary phase at 19 °C. Cellular extracts for mitochondrial proteins were analysed. The changes in protein abundance in *mia40-4int*[S] were as expected for the *mia40-4int* strain. **k**, Proteins quantified in three independent biological replicates were plotted according to their *P* value (log10) against the log10-transformed *mia40-4int*[S]/WT[S] ratios. Proteins with a *P* value <0.05 and a fold-change in protein abundance >1.5 or <−1.5 were considered upregulated and downregulated and are marked in green and red, respectively. The MIA pathway substrates are highlighted by enlarged circles. WT, wild-type; p, precursor; m, mature; asterisk indicates unspecific band; IAA, iodoacetamide. **f, g, i, j**, Uncropped blots/gels are in Supplementary Information Fig. 1.
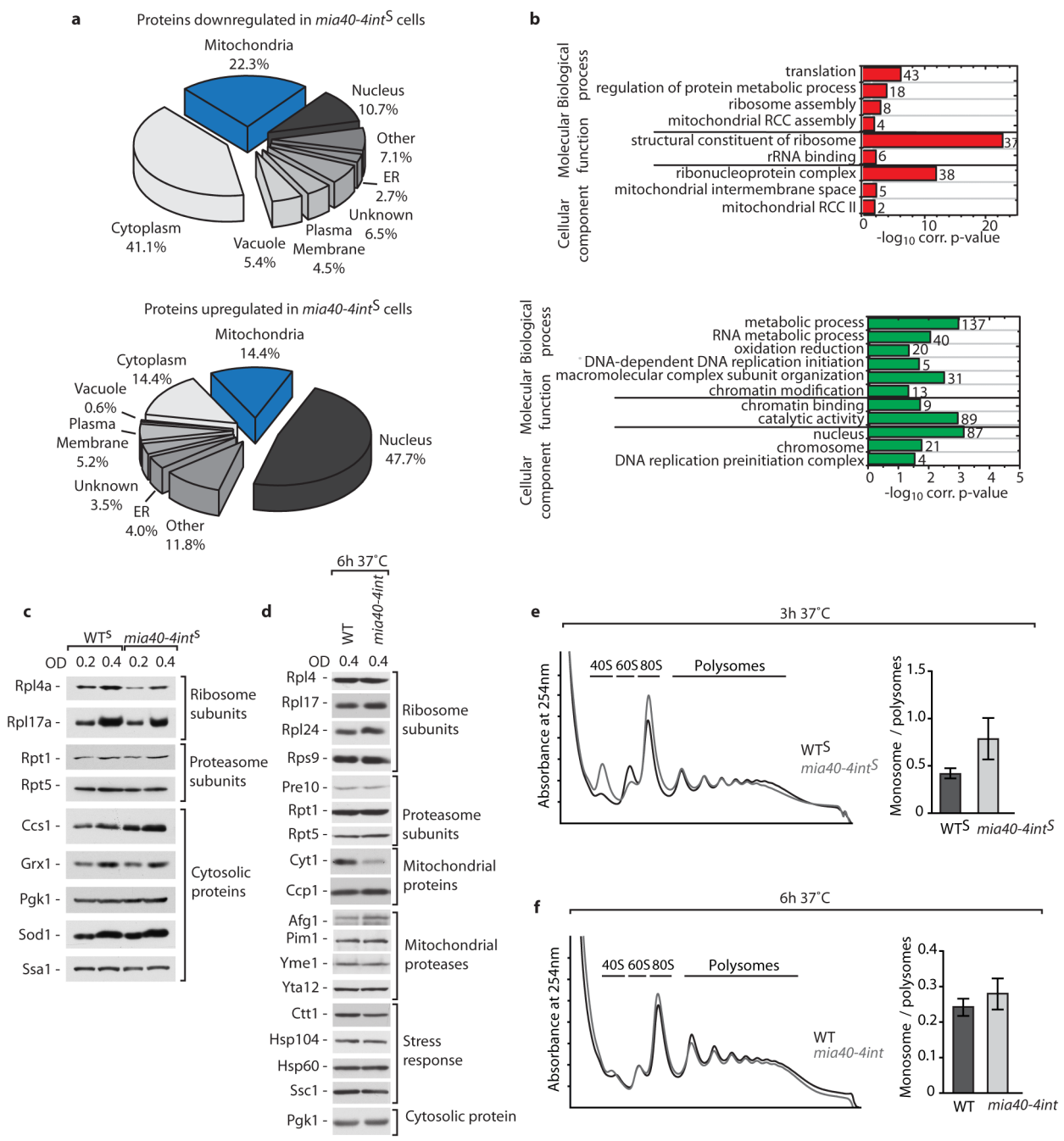
Extended Data Figure 2 | Protein abundance in *mia40-4int* and *mia40-4int^s*. **a**, Distribution of proteins quantified by SILAC-MS analysis of WT^S and *mia40-4int^S* cells based on the GO terms for cellular components provided by the *Saccharomyces* genome database. Subcellular localizations are shown for fractions of proteins with a significant 1.5-fold change in abundance and a *P* value < 0.05 in the *mia40-4int^S* cells. **b**, GO term enrichment analysis of proteins found to be significantly downregulated (top) or upregulated (bottom) in *mia40-4int^S*. **c**, WT^S and *mia40-4int^S* yeast were grown to the stationary phase at 19 °C. Cellular extracts were analysed for non-mitochondrial proteins. **d**, WT and *mia40-4int* yeast were grown in the respiratory medium to the

logarithmic phase at 19 °C and shifted for 6 h to 37 °C. Cellular extracts were analysed. No changes in protein abundance in *mia40-4int* in comparison to wild type were observed. **c, d**, Uncropped blots are in Supplementary Information Fig. 1. **e, f**, Yeast were cultured in the full medium with galactose to early logarithmic phase and shifted for 3 h to 37 °C (*mia40-4int^s*) (**e**) or 6 h 37 °C (*mia40-4int*) (**f**). Cells were treated with 100 μg ml^{-1} cycloheximide for 10 min. Yeast lysates were fractionated on a 10–50% linear sucrose gradient and absorbance was monitored at 254 nm. The retention of 40S, 60S ribosomal subunits, monosomes (80S) and polysomes is indicated. The monosomes versus polysomes ratio was quantified. Mean ± s.e.m., *n* = 3. WT, wild type.
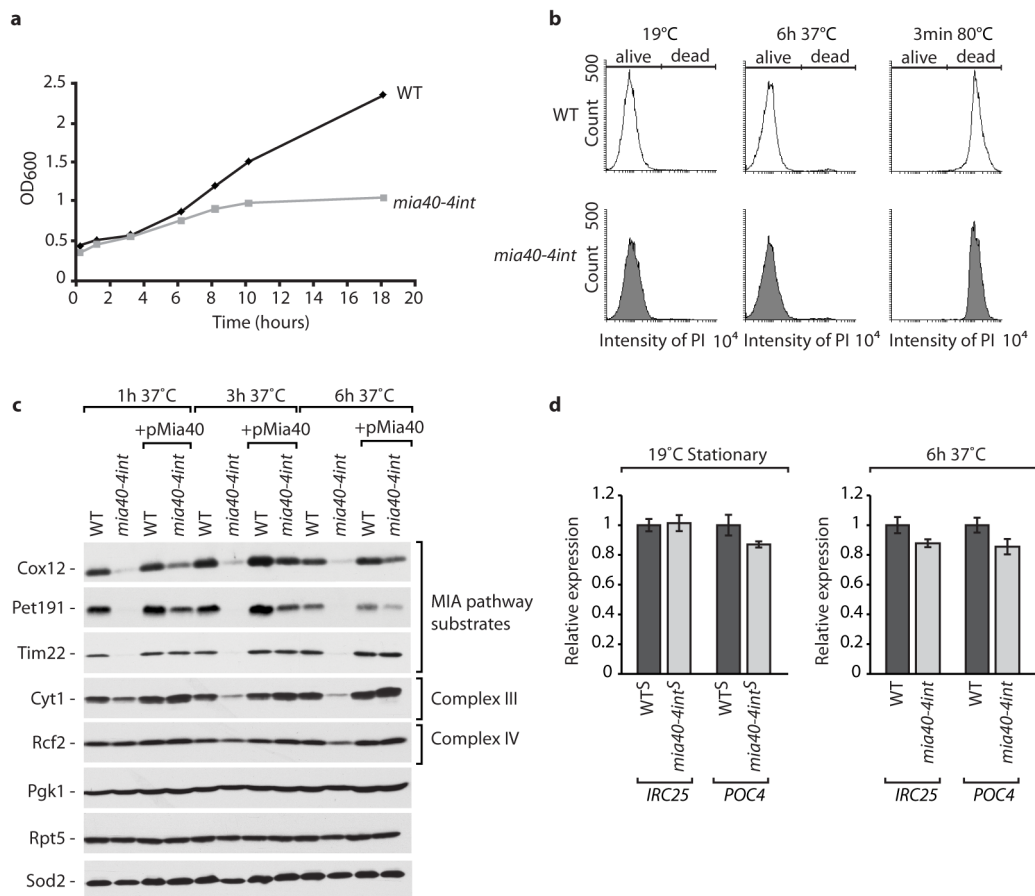
Extended Data Figure 3 | Characterization of the *mia40-4int* mutant.
a, Wild type and *mia40-4int* were grown in respiratory medium and their growth was compared upon shift to the restrictive temperature of 37 °C. D_{600nm} (OD600) was measured at the indicated time points. **b**, The survival of wild type and *mia40-int* grown to the logarithmic phase at 19 °C, heat-shocked for 6 h at 37 °C or heat-killed for 3 min at 80 °C was assessed by propidium iodide (PI) staining. **c**, The *mia40-4int* and wild-type cells transformed or not transformed with the plasmid encoding Mia40 were grown in respiratory medium at 19 °C and shifted to 37 °C for the indicated times. Protein levels were analysed. The *mia40-4int*-dependent defect in protein levels was complemented by *MIA40*. Uncropped blots are in Supplementary Information Fig. 1. **d**, mRNA levels of *IRC25* or *POC4* in *mia40-4int*[S], *mia40-4int* and the corresponding wild type. Cultures were grown in respiratory medium at 19 °C. The heat stress was conducted at 37 °C for the indicated time. Transcript levels of *ALG9*, *FBA1* and *TUB2* were used for normalization. Wild type was set to 1. Mean ± s.e.m., $n = 3$. WT, wild type.
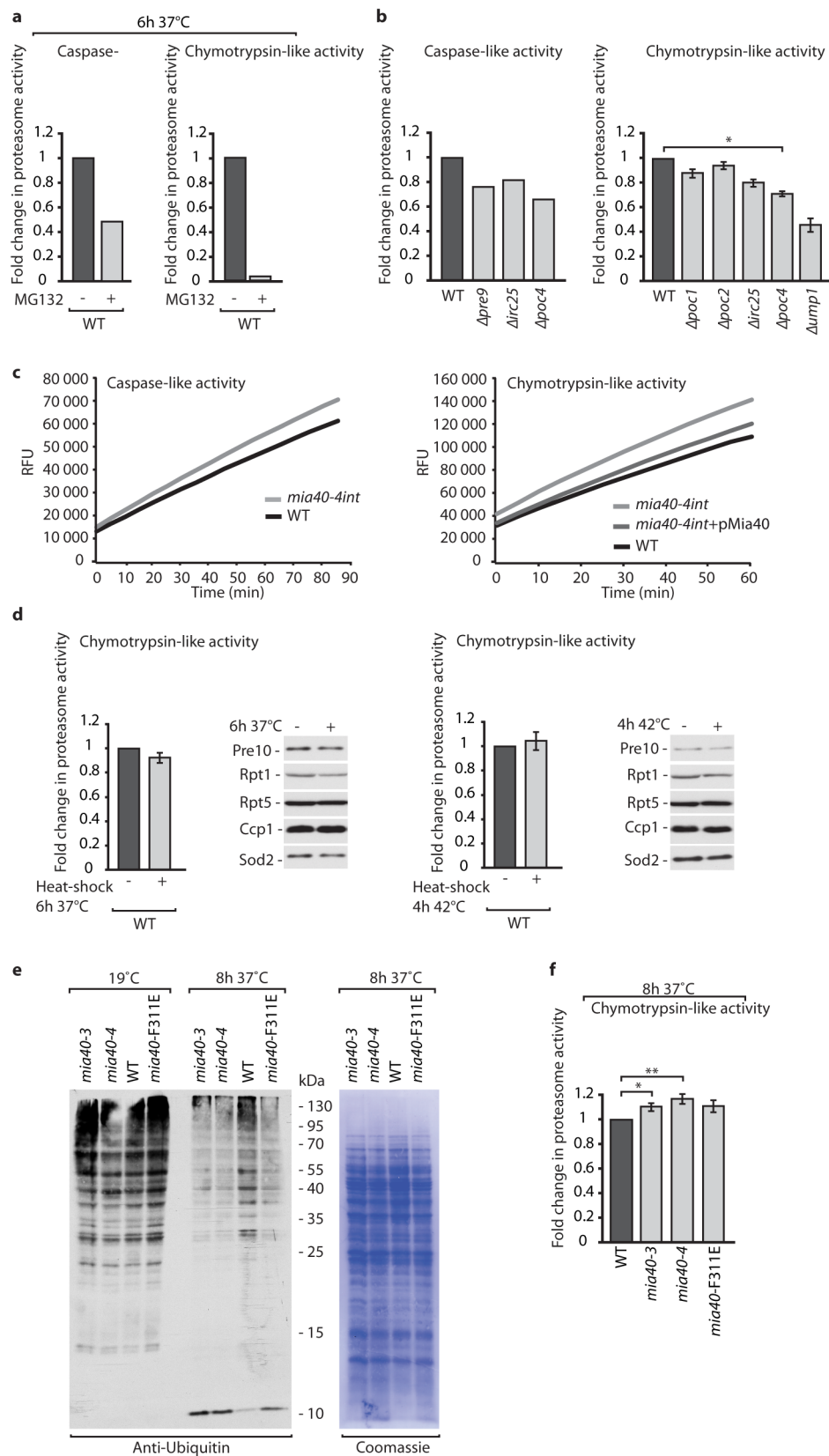
Extended Data Figure 4 | Proteasomal activity in *mia40* mutants. **a**, Wild-type cells were grown in respiratory medium at 19 °C and shifted to 37 °C for 6 h. Where indicated, cell lysates were incubated with 50 μM MG132 for 2 min before addition of caspase-like or chymotrypsin-like proteasome substrate. The activity was inhibited upon MG132 addition, confirming the proteasomal specificity of the assay. **b**, Wild-type cells and cells deleted for *POC1, POC2, IRC25, POC4, UMP1* and *PRE9* were grown in respiratory medium to stationary phase at 24 °C. The proteasome caspase-like and chymotrypsin-like activities were reduced upon the compromised proteasome demonstrating the specificity of the assay. Mean ± s.e.m., $n = 3$. *P value < 0.05. **c**, Cells were grown in respiratory medium to logarithmic phase at 19 °C and shifted to 37 °C for 6 h. Proteasomal activities were analysed over time. Left, mean of 7 biological replicates; right, mean of 4 biological replicates. **d**, Wild-type cells were grown at 24 °C and were shifted for 6 h at 37 °C or for 4 h at 42 °C. Proteasome activities and the levels of proteasomal subunits, Pre10, Rpt1 and Rpt5 were not changed. Mean ± s.e.m., $n = 3$. **e**, Total protein extracts of plasmid-borne *mia40* mutants were analysed by anti-ubiquitin immunoblotting and Coomassie staining. **d**, **e**, Uncropped blots/gels are in Supplementary Information Fig. 1. **f**, Proteasomal activity of wild type and *mia40* mutants. Mean ± s.e.m., $n = 4$. *$P < 0.05$; **$P < 0.03$. WT, wild type; RFU, relative fluorescent units.
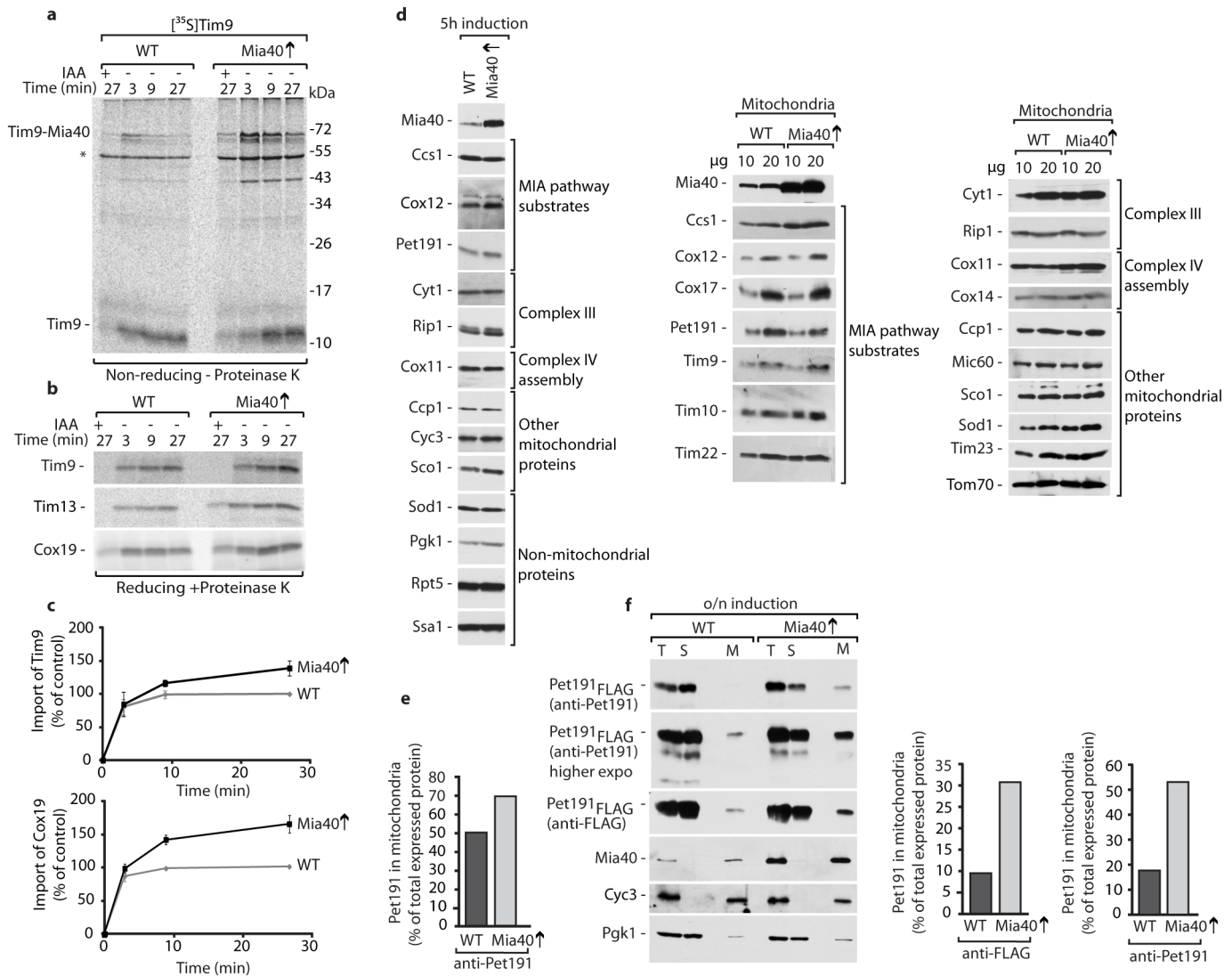
Extended Data Figure 5 | Characterization of cells that overproduce Mia40.
a, Formation of a disulfide-bonded intermediate between Mia40 and Tim9 is accelerated in mitochondria with overproduced Mia40. Mitochondria were isolated from wild-type and Mia40 overproducing (Mia40↑) strains and incubated with [^{35}S]-labelled Tim9 precursor. When indicated, iodoacetamide (IAA) was added as a control to block mitochondrial import. **b**, More efficient import of proteins in mitochondria with overproduced Mia40. Mitochondria from WT and Mia40↑ strains were incubated with [^{35}S]-labelled precursors. The samples were treated with proteinase K to remove non-imported proteins. **c**, Quantification of [^{35}S]radiolabelled Tim9 and Cox19 import. Mean ± s.e.m.,

$n = 3$. **d**, Cellular and mitochondrial protein levels were analysed in wild-type and Mia40↑ strains by western blotting. The overproduction of Mia40 did not change the protein levels. **e**, Quantification of Pet191 in mitochondria (M) in wild type and Mia40↑ after 6 h induction (Fig. 1i). **f**, The WT and Mia40↑ cells producing Pet191$_{Flag}$ were grown in fermentable medium with 2% glucose at 24 °C and shifted to galactose-containing medium for overnight induction. Protein levels in total (T), post-mitochondrial supernatant (S) and mitochondria (M) were analysed. The mitochondrial localization (M) of Pet191 in WT and Mia40↑ after overnight induction was quantified. WT, wild type. **a, b, d, f**, Uncropped gels/blots are in Supplementary Information Fig. 1.
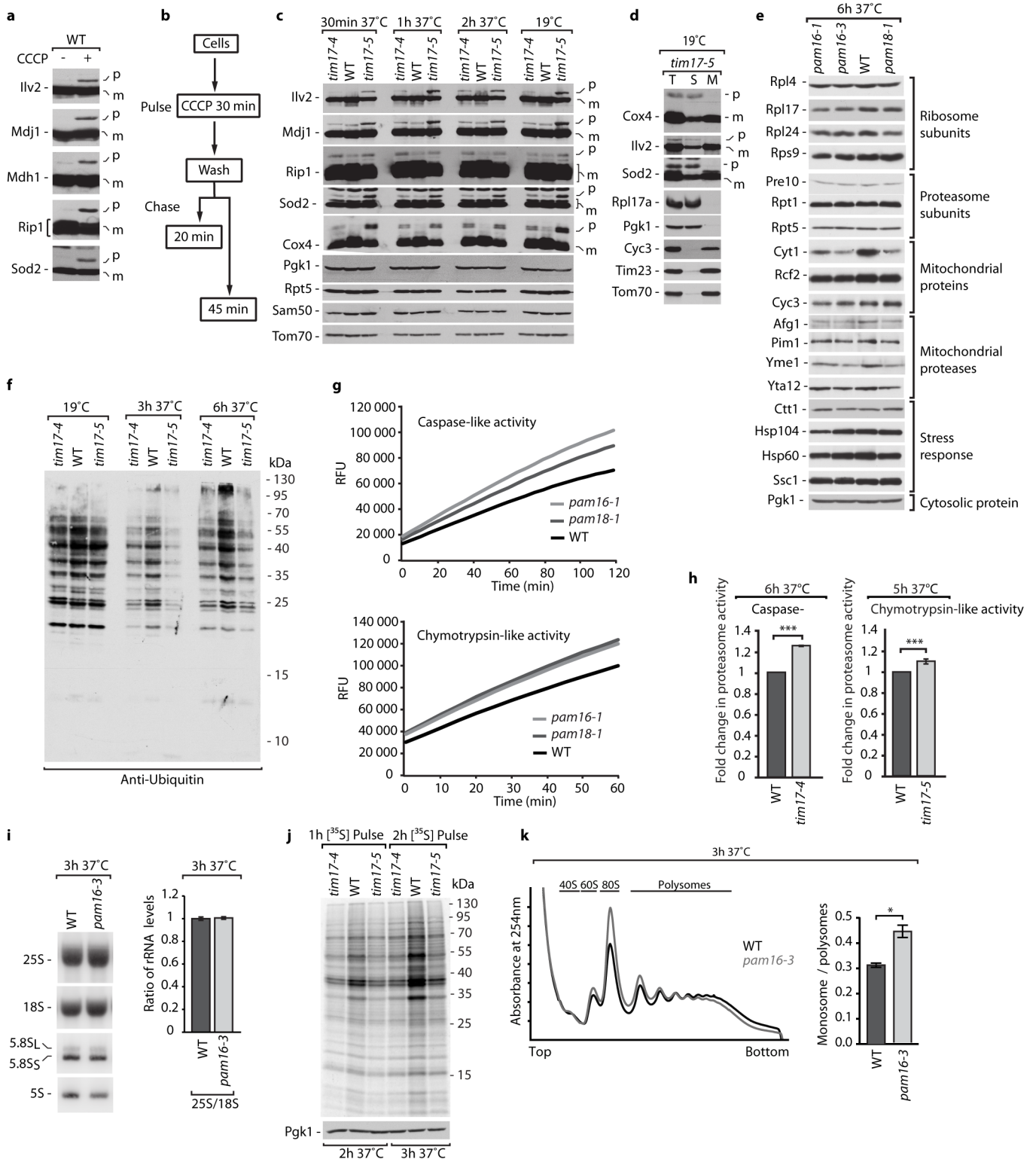
**Extended Data Figure 6 | Characterization of mutants that affect the import via the TIM23 presequence pathway. a**, The wild-type strain was grown to stationary phase at 19 °C and treated with CCCP for 2 h to dissipate the electrochemical potential of the inner mitochondrial membrane. The CCCP treatment resulted in the accumulation of precursor proteins. **b**, Representation of the pulse-chase experiment. The wild type was treated for 30 min with CCCP (pulse) and chased in fresh medium without CCCP for 20 or 45 min for analysis. **c**, Non-processed precursor proteins accumulate in the *tim17* mutants. The *tim17* mutants and the corresponding wild type were grown in fermentative medium to stationary phase at 19 °C, shifted to 37 °C and analysed by western blotting. **d**, The *tim17-5* mutant was grown to stationary phase at 19 °C and shifted to 37 °C for 90 min. The cells were fractionated and equal volumes of total (T), post-mitochondrial supernatant (S) and mitochondrial (M) fractions were analysed by western blotting. Precursor proteins were localized in the cytosol together with cytosolic proteins (Rpl17 and Pgk1), in contrast to mature mitochondrial proteins (Cyc3, Tim23, Tom70). **e**, The *pam16* and *pam18* mutants were grown to logarithmic growth phase and shifted to restrictive temperature. Total protein content was analysed by western blotting. No changes in protein levels were detected with the exception for a small decrease in the ribosomal proteins Rpl17 and Rpl24 was noticed. **f**, Ubiquitinated species decreased in the *tim17-4 and tim17-5* mutants. The *tim17* mutants and corresponding WT strain were grown in respiratory medium at 19 °C, shifted to 37 °C and analysed by anti-ubiquitin immunoblotting. **g**, The *pam16-1* and *pam18-1* mutants were grown in respiratory medium at 19 °C and shifted to 37 °C for proteasomal activity assays. Mean of 3 biological replicates. **h**, The *tim17* mutants and corresponding wild-type strain were grown in respiratory medium at 19 °C and shifted to 37 °C for proteasomal activity assays. Mean ± s.e.m., $n = 3$ (caspase-like activity), $n = 5$ (chymotrypsin-like activity). ***$P < 0.01$. **i**, Northern blot of rRNA and quantification. The *pam16-3* and wild-type cells were grown in respiratory medium and shifted to 37 °C. Mean ± s.e.m., $n = 3$. **j**, Incorporation of [35S]-labelled amino acids is decreased in *tim17* mutants compared to wild type. Strains were grown in respiratory medium and shifted to 37 °C. Samples were taken after 1 or 2 h of [35S] labelling and analysed by SDS–PAGE and autoradiography. **k**, Representative gradient profiles of ribosomes in *pam16-3* and wild type and quantification of the monosome versus polysome fractions. Mean ± s.e.m., $n = 3$. *$P < 0.05$. Cells were grown to logarithmic phase, shifted to 37 °C for 3 h and treated with 100 µg ml^{-1} cycloheximide for 10 min. Lysates were fractionated on a 10–50% linear sucrose gradient and absorbance was monitored at 254 nm. The monosomes versus polysomes ratio was quantified. WT, wild-type. RFU, relative fluorescent units. **a**, **c–f**, **i**, **j**, Uncropped blots/gels are in Supplementary Information Fig. 1.
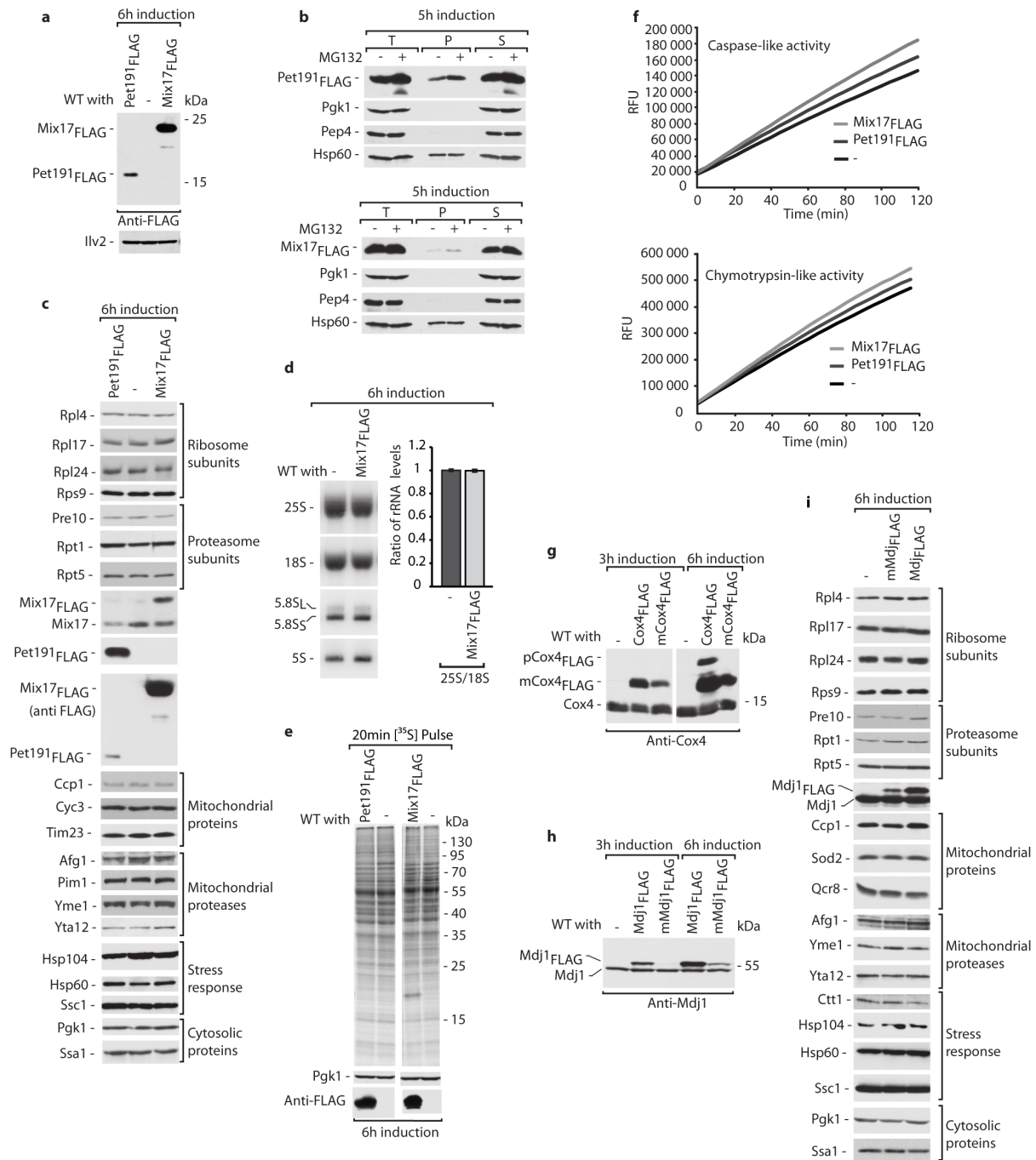
Extended Data Figure 7 | Translation and proteasomal activity in the cells overproducing mitochondrial proteins. **a**, Pet191$_{Flag}$ and Mix17$_{Flag}$ were expressed in WT cells at 24 °C. **b**, Flag-tagged proteins were expressed in WT and when indicated cells were treated with MG132 for 4 h. Fractions of total (T), aggregates (P) and soluble (S) proteins were analysed by western blotting. **c**, Wild type expressing Pet191$_{Flag}$ or Mix17$_{Flag}$ were grown at 24 °C and analysed for total protein content by western blotting. No changes in protein levels compared to wild type were found, including ribosomal or proteasome subunits. **d**, Northern blot analysis and quantification of rRNA in cells expressing Mix17$_{Flag}$. Mean ± s.e.m., $n = 3$. **e**, Mistargeted mitochondrial proteins do not alter the rate of translation. Incorporation of [^{35}S]-labelled amino acids in wild type expressing Pet191$_{Flag}$ or Mix17$_{Flag}$. The expression of

Flag-tagged proteins was induced for 6 h at 24 °C. Samples were taken 20 min after initiation of [^{35}S] labelling and analysed by SDS–PAGE and audioradiography. **f**, Expression of Pet191$_{Flag}$ or Mix17$_{Flag}$ stimulates proteasomal activity. Mean of 3 biological replicates. **g**, Cox4$_{Flag}$ with (pCox4$_{FLAG}$) or without mitochondrial presequence (mCox4$_{Flag}$) was expressed in wild type. **h**, Mdj1$_{Flag}$ and mMdj1$_{Flag}$ proteins were expressed in WT cells. The presence of Flag-tagged proteins was confirmed by immunoblotting. **i**, Wild type expressing Mdj1$_{Flag}$ or mMdj1$_{Flag}$ were grown at 24 °C and analysed for total protein levels by western blotting. No changes in protein levels compared to control were found, including ribosomal and proteasome subunits. WT, wild type. RFU, relative fluorescent units. **a–e, g–i**, Uncropped blots are in Supplementary Information Fig. 1.
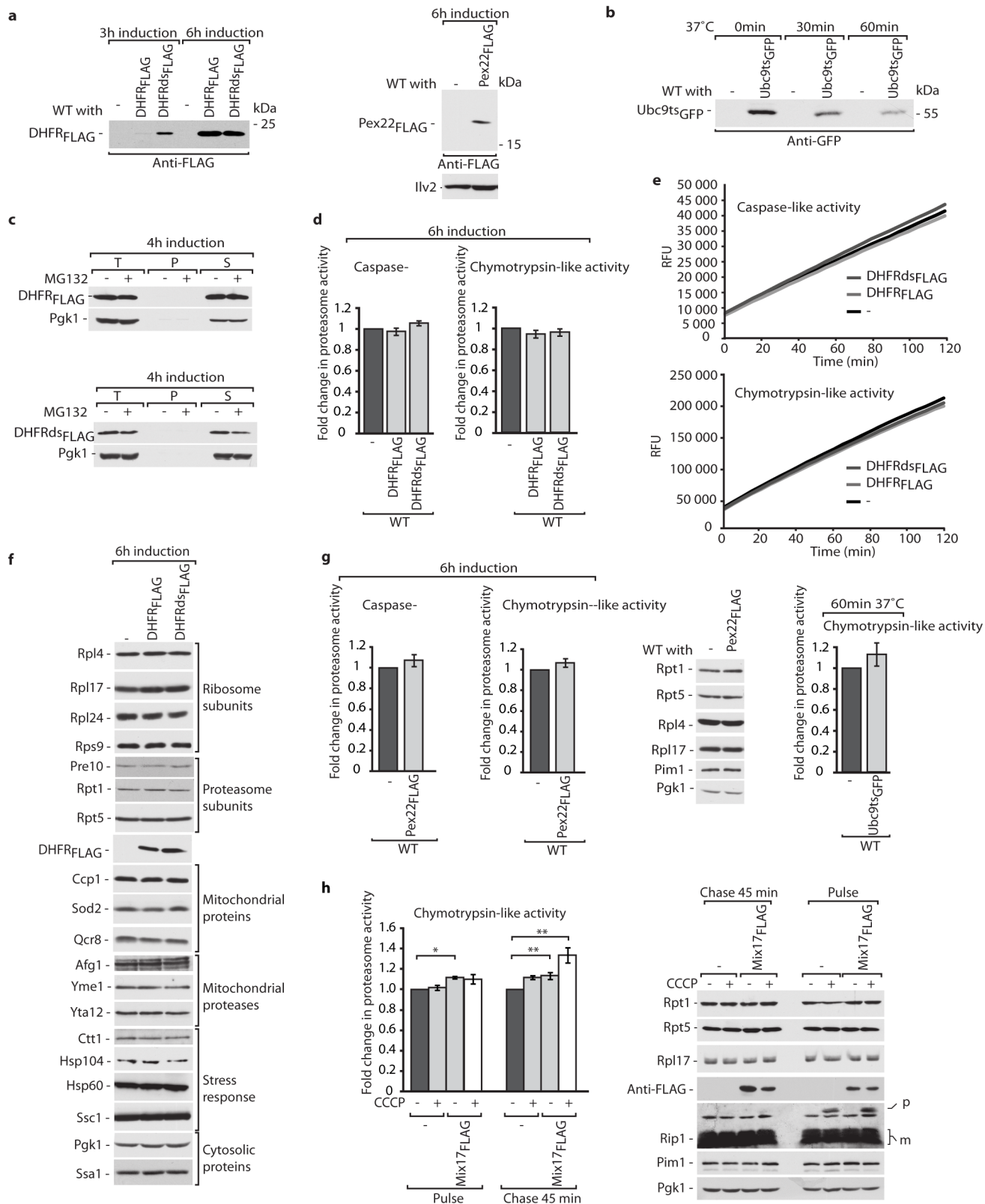
Extended Data Figure 8 | Proteasomal activity in the cells overproducing non-mitochondrial proteins. a, Expression of DHFR$_{Flag}$ and DHFRds$_{Flag}$ was induced at 24 °C and Pex22$_{Flag}$ at 28 °C in wild type. The presence of Flag-tagged proteins was confirmed by immunoblotting. **b**, Ubc9ts–GFP was induced in galactose and subsequently wild-type cells were shifted to glucose medium at 37 °C to initiate unfolding. After indicated time points, samples were analysed by western blotting. **c**, Flag-tagged proteins were expressed in wild type for the indicated time and the proteasome was inhibited with MG132 for 3 h if indicated. Fractions of total (T), aggregates (P) and soluble (S) proteins were analysed and no aggregation was observed. **d**, **e**, Proteasomal activity in WT expressing DHFR$_{Flag}$ or DHFRds$_{Flag}$ grown at 24 °C. Mean ± s.e.m. $n = 6$ (**d**). Mean of 6 biological replicates (**e**). **f**, Wild-type cells expressing DHFR$_{Flag}$ or DHFRds$_{Flag}$ were grown at 24 °C and analysed for total protein content. No changes in protein levels were found compared to wild type, including ribosomal and proteasome subunits. **g**, Proteasomal activity in wild type expressing Pex22$_{Flag}$ grown at 28 °C and the abundance of proteins were not significantly changed. No significant change in proteasomal activity was detected upon expression of Ubc9ts–GFP. Mean ± s.e.m., $n = 6$. **h**, The proteasomal stimulation by CCCP and Mix17$_{Flag}$ is additive. Wild-type cells overproducing Mix17$_{Flag}$ were treated with CCCP to measure the chymotrypsin-like activity of the proteasome. In the case of Mix17$_{Flag}$, proteasomal stimulation was less efficient than stimulation reported for Mix17$_{Flag}$ in Fig. 3a due to change in experimental conditions imposed by the CCCP treatment. Mean ± s.e.m., $n = 3$. *$P < 0.05$; **$P < 0.03$. Analysis of cellular protein content showed no difference in proteasome subunits (Rpt1, Rpt5) or ribosomal protein Rpl17. WT, wild type. RFU, relative fluorescent units. **a–c**, **f–h**, Uncropped blots are in Supplementary Information Fig. 1.
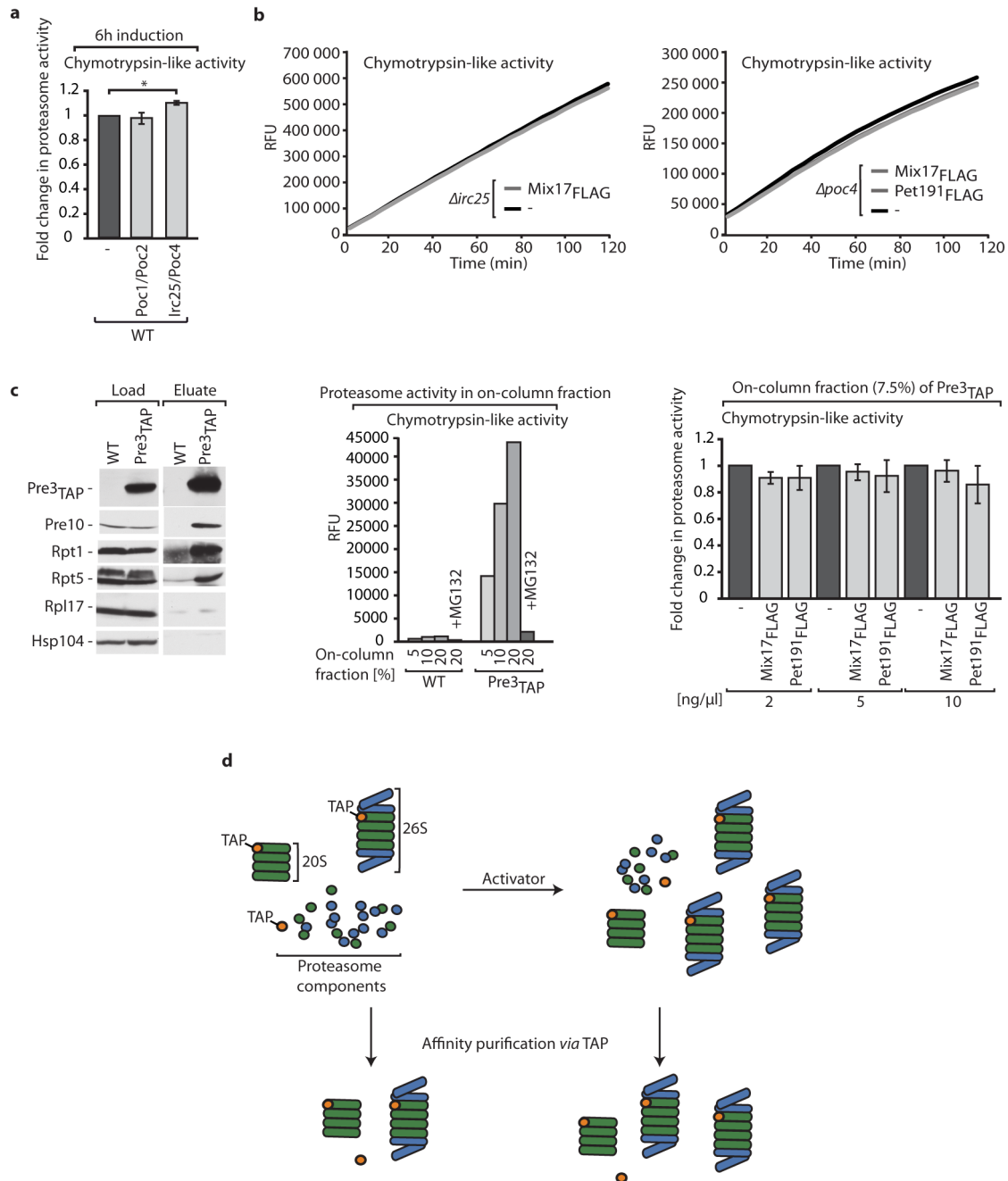
Extended Data Figure 9 | Auxiliary factors are required to stimulate proteasome by mitochondrial precursor proteins. a, Proteasomal activity in wild-type cells expressing simultaneously either Poc1$_{Flag}$ and Poc2$_{Myc}$ or Irc25$_{Flag}$ and Poc4$_{Myc}$. Mean $\pm$ s.e.m., $n = 3$. *$P < 0.05$. The overexpression of *POC1* and *POC2* or *IRC25* and *POC4* was induced from one plasmid. The overexpression of Irc25$_{Flag}$ and Poc4$_{Myc}$ led to a small increase in the proteasomal activity, in spite of the inability to detect these proteins likely due to tight regulation of their abundance (not shown). **b**, Proteasomal activity in cells lacking Irc25 or Poc4, expressing Mix17$_{Flag}$ or Pet191$_{FLAG}$ and grown at 24 °C. Mean of 4 biological replicates. **c**, Affinity purification of the proteasome complex via Pre3$_{TAP}$ from cells grown at 28 °C. Load, 5%; eluate, 100%. Chymotrypsin-like activity of the proteasome bound to the column was measured. The specificity was checked by the treatment of the on-column

fraction with proteasomal inhibitor MG132. Activity of the purified proteasome via Pre3$_{TAP}$ was measured upon addition of purified Mix17$_{Flag}$ or Pet191$_{Flag}$ (for 7.5% of the on-column fraction). Mean $\pm$ s.e.m., $n = 3$. Uncropped blots are in Supplementary Information Fig. 1. **d**, Representation of the proteasome complex affinity purification. The subunits of proteasome are assembled into the 20S catalytic core and the 19S regulatory particle. The core and regulatory particles joined together to form the 26S proteasome. Overexpression of an activator (that is, Mix17) stimulates the 26S proteasome assembly. Thus, upon affinity purification via a TAP-tagged proteasome subunit, more proteasomal subunits representing more assembled proteasomes are found in the eluate in the presence of an activator. WT, wild type. RFU, relative fluorescent units.
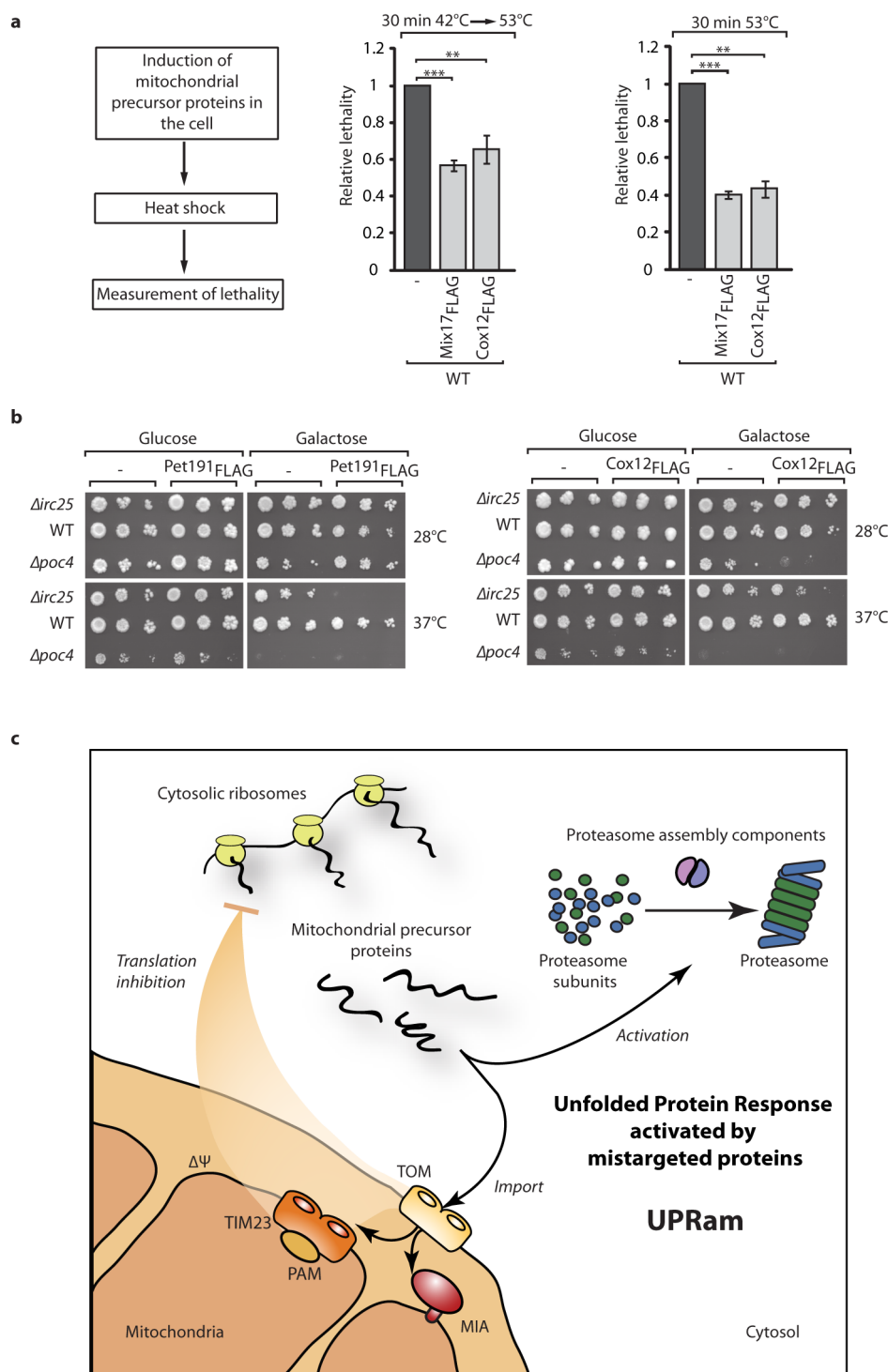
Extended Data Figure 10 | The proteasome assembly heterodimer Irc25–Poc4 is required to protect cells against stress. a, Representation of heat stress experiments. Cells overexpressing mitochondrial proteins were exposed to different heat shock conditions and subsequently subjected to lethality assessment (left panel). Cells were exposed to a gradual increase in temperature from 42 °C to 53 °C within 30 min or were incubated at 53 °C for 30 min. Mean lethality values of wild-type cells expressing empty plasmid increased with harsher stress conditions (9% for middle panel; 22% for right panel). The lethality of cells expressing empty plasmid was set to 1. Mean ± s.e.m., $n = 5$ (middle panel), $n = 4$ (right panel). **$P < 0.03$; ***$P < 0.01$. **b**, Wild type or cells deleted for the *IRC25 or POC4* genes and overproducing Pet191$_{Flag}$ or Cox12$_{Flag}$ protein were cultured on agar plates with sucrose. Consecutive tenfold dilutions of cells were spotted on selective medium plates with either glucose or galactose. Cells were grown at the indicated temperatures. **c**, Model for cellular responses activated by the mitochondrial protein import and precursor over-accumulation stress. WT, wild type.

LETTER

Kinetochore-localized PP1–Sds22 couples chromosome segregation to polar relaxation

Nelio T. L. Rodrigues[1], Sergey Lekomtsev[1], Silvana Jananji[2], Janos Kriston-Vizi[1], Gilles R. X. Hickson[2,3] & Buzz Baum[1,4,5]

Cell division requires the precise coordination of chromosome segregation and cytokinesis. This coordination is achieved by the recruitment of an actomyosin regulator, Ect2, to overlapping microtubules at the centre of the elongating anaphase spindle[1]. Ect2 then signals to the overlying cortex to promote the assembly and constriction of an actomyosin ring between segregating chromosomes[1]. Here, by studying division in proliferating *Drosophila* and human cells, we demonstrate the existence of a second, parallel signalling pathway, which triggers the relaxation of the polar cell cortex at mid anaphase. This is independent of furrow formation, centrosomes and microtubules and, instead, depends on PP1 phosphatase and its regulatory subunit Sds22

(refs 2, 3). As separating chromosomes move towards the polar cortex at mid anaphase, kinetochore-localized PP1–Sds22 helps to break cortical symmetry by inducing the dephosphorylation and inactivation of ezrin/radixin/moesin proteins at cell poles. This promotes local softening of the cortex[2,3], facilitating anaphase elongation and orderly cell division. In summary, this identifies a conserved kinetochore-based phosphatase signal and substrate, which function together to link anaphase chromosome movements to cortical polarization, thereby coupling chromosome segregation to cell division.

As animal cells exit mitosis they assemble a cleavage furrow at their equator, whose position is determined by the overlapping microtubules of the central spindle[1]. This ensures the precise coordination
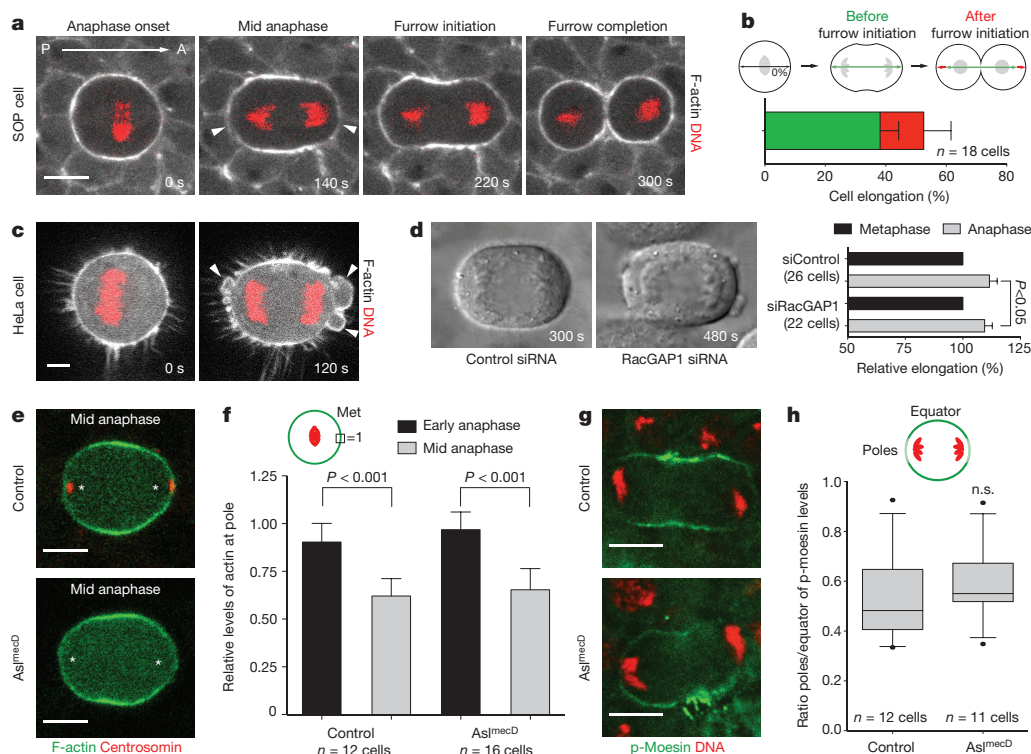

Figure 1 | Polar relaxation and anaphase cell elongation are independent of furrow ingression, centrosomes and astral microtubules. a, b, Representative stills and graph to show SOP cell elongation during mitotic exit, before and after furrowing (18 cells). A–P denotes anterior–posterior axis. **c,** Equivalent HeLa cell imaged at mitotic exit (representative of 10 biological replicates). **d** (left and right), Representative DIC stills for 26 control and 22 RacGAP1-depleted anaphase cells, together with quantification of cell elongation in each case relative to metaphase (the difference is small but significant, *P* = 0.038). **e, f,** Levels of polar F-actin in early and mid anaphase in 12 control and 16 Asl^mecD SOP cells (**e** shows representative images of mid-anaphase cells

quantified in **f**). F-actin is labelled with Lifeact–GFP in **a** and **c**, and by GMA in **e**. Asterisks in **e** mark chromosomes. Met, metaphase. **g, h,** Poles/equator ratio of cortical p-moesin levels in immunostained fly cells (**g** shows representative image of cells quantified in **h**; n.s., not significant (*P* = 0.388)). DNA labelled with Hoechst 33258 in **g**. White arrowheads in **a** and **c** indicate polar relaxation. Anaphase onset = 0 s. Scale bars, 5 μm. Bar graphs show mean and s.d. box-and-whisker plots show median and 10–90 percentiles. Significance in each case was assessed using a two-tailed unpaired *t*-test. DIC, differential interference contrast; GFP, green fluorescent protein; GMA, GFP fused to the F-actin-binding domain of *Drosophila* moesin.

[1]MRC Laboratory for Molecular Cell Biology, University College London, Gower Street, London WC1E 6BT, UK. [2]Sainte-Justine Hospital Research Center, Montréal, Québec H3T 1C5, Canada. [3]Department of Pathology and Cell Biology, Université de Montréal, Montréal, Québec H3T 1J4, Canada. [4]Institute for the Physics of Living Systems, University College London, Gower Street, London WC1E 6BT, UK. [5]CelTisPhyBio Labex, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France.

of cell division and chromosome segregation. Recently, however, it has become clear that proliferating animal cells assemble a rigid and isotropic actomyosin cortex as they enter mitosis[4], which, if not disassembled, might interfere with furrow formation. To test this idea, and to determine whether relaxation of the cortex at opposing cell poles contributes to cell division[5], we analysed sensory organ precursor (SOP) cells in the *Drosophila* pupal notum, where live-cell imaging can be used to study the changes in cell shape that accompany mitotic progression in the context of a developing epithelium[6,7]. In these cells, actin was cleared from cell poles before any overt signs of equatorial constriction were visible (Fig. 1a, b), as previously proposed[8]. Similarly, polar relaxation was found to precede furrow formation in human cells (Fig. 1c). Moreover, in both systems, polar relaxation, as evidenced by actin clearance, cell elongation and/or blebbing, was seen following silencing of the centralspindlin component RacGAP1 (also called MgcRacGAP in humans and Tumbleweed in *Drosophila*) (Fig. 1d and Extended Data Figs 1a, d and 2a, b), although this led to profound defects in the equatorial recruitment of actin and myosin II, furrow formation and cytokinesis (Extended Data Figs 1a–c, e and 2c, d) (following RacGAP1/Tumbleweed depletion, 100% of human cells failed to divide[9,10] as did 30% of *Drosophila* cells). In addition, actin (Extended Data Fig. 2e–g) and phospho-T559-moesin (p-moesin) (Extended Data Fig. 2h) disappeared from the polar cortex before their accumulation at the furrow. Together, these data suggest the existence of a conserved second signal that functions independently of the well-studied spindle midzone to polarize the anaphase cortex.

To identify the source of the signal, we began by considering the possibility that it is associated with centrosomes and astral microtubules, which have previously been proposed to regulate cortical contractility at the cell equator[11,12] and at cell poles[13,14]. We eliminated both through mutation of the gene *asterless* (ref. 15) (Extended Data Fig. 3a–c). This led to randomization of the orientation of the spindle (Extended Data Fig. 3d) without compromising mitotic progression, as expected following a loss of crosstalk between the metaphase spindle and cortex. Nevertheless, in anaphase, both F-actin (Fig. 1e, f and Extended Data Fig. 3e) and p-moesin (Fig. 1g, h) were cleared from opposing cell poles as efficiently as they were in control cells.

As an alternative approach to address the same question, we used a cyclin-dependent kinase (CDK) inhibitor (RO3306) to force cultured cells arrested in prometaphase to leave mitosis with a monopolar spindle. This induces monopolar cytokinesis[16]. As cortical symmetry was broken, phosphorylated ezrin/radixin/moesin (p-ERM) was lost from the cortex on the side of the cell furthest from centrosomes and the bulk of midzone microtubules (Extended Data Fig. 3f). Similar results were observed in cells forced to exit mitosis in the presence of high doses of microtubule inhibitors following either CDK inhibition (Extended Data Fig. 3g–i, k) or Mad2 depletion (Extended Data Fig. 3g, h, j, k). These data show that the cortical polarization induced by mitotic exit is independent of both centrosomes and astral microtubules. Notably, however, actin filaments and p-ERM proteins were always lost from portions of the cortex closest to anaphase chromatin in these experiments (Extended Data Fig. 3f–k). This observation suggests a role for anaphase chromatin in the establishment of cortical polarity. To assess the timing of this chromosomally associated signal, we repeated the analysis in cells forced to flatten through the expression of an activated version of the Rap1 GTPase (Rap1*)[17]. Although this pushed metaphase chromosomes close to the cell cortex, forced mitotic exit (through CDK inhibition) still triggered a sudden loss of F-actin and cortical relaxation close to chromatin (Extended Data Fig. 4a, b). Thus, the chromatin-associated signal appears enhanced at the onset of anaphase.

To see whether segregating chromosomes perform a similar function during a normal mitotic exit (Fig. 2a and Supplementary Video 1), we used kymographs of dividing SOP cells (Fig. 2b) to monitor the relative timing of anaphase cell elongation (D1), the distance of the chromatin (kinetochore interface) to the cell centre (D2), and the approach of chromosomes to the poles (D3) (Fig. 2c and Extended
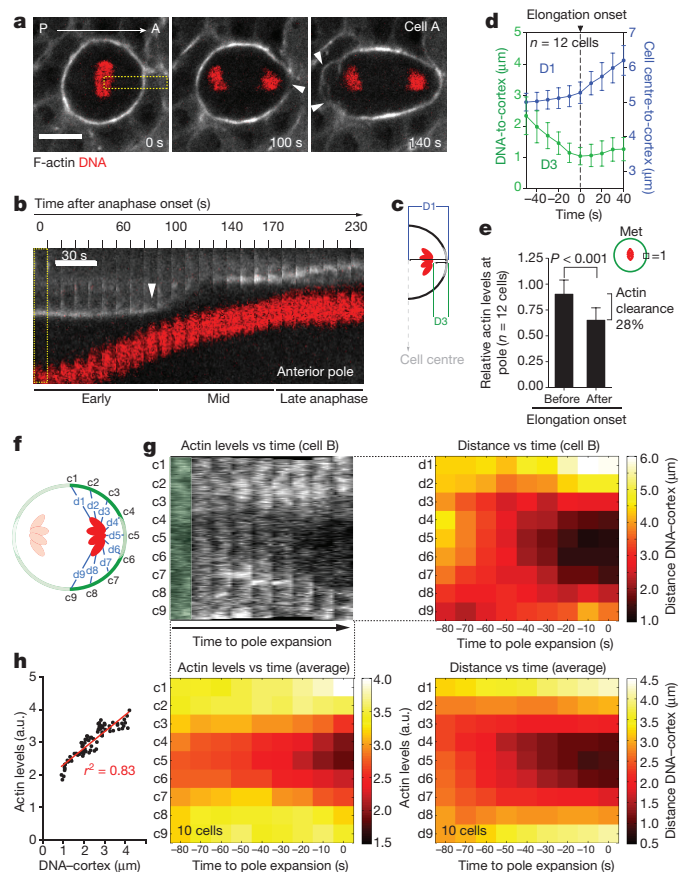

Figure 2 | Polar relaxation is triggered by chromatin proximity in mid anaphase. **a**, **b**, Images and the corresponding kymograph for an SOP cell undergoing anaphase (representative of 12 biological replicates). In **a**, arrowheads indicate polar relaxation. Scale bar, 5 μm. Anaphase onset = 0 s. Yellow box denotes cross-section. **c**, Scheme depicting distances D1 and D3. **d**, **e**, Show distances D1 and D3 (**d**) and relative actin levels (**e**) before and after the onset of elongation. **f**, Scheme of cortical regions c1–c9 and DNA-to-cortex distances d1–d9. **g**, Kymographs of cortical actin and DNA-to-cortex distance over time for a representative SOP cell (top), and when averaged over 10 cells (bottom). **h**, Actin levels (seen in **g**, bottom left) were strongly correlated ($r^2 = 0.83$) with DNA-to-cortex distance (seen in **g**, bottom right). Actin levels expressed in arbitrary units (a.u.). F-actin was labelled with Lifeact–GFP. Graphs show mean with s.d. Significance was assessed using a two-tailed unpaired *t*-test.

Data Fig. 4c, d). Under these physiological conditions the approach of the chromatin mass to cell poles was associated with loss of cortical actin (Fig. 2d, e), which translated into a simple, near-linear relationship between the distance of kinetochores from the cell pole and cortical actin clearance (Fig. 2f–h). Furthermore, in SOP cells, cortical actin was lost from the anterior pole, where chromatin approaches first, before being lost from the posterior pole (Extended Data Fig. 4f–i). Taken together, these data support the existence of a chromosomally derived signal that induces polar relaxation following mitotic exit.

PP1-87B phosphatase and its regulatory subunit Sds22 (also called PPP1R7 in humans) seemed good candidates for molecular regulators of this process as previous work showed that they negatively regulate p-moesin during mitotic exit[2,3]. As previously reported[2], RNA interference (RNAi)-mediated silencing of either PP1-87B or Sds22 led to a failure in clearance of the polar cortex at anaphase (Fig. 3a, b and Supplementary Videos 2, 3). This was accompanied by a defect in cell elongation (Extended Data Fig. 5c–h), enabling anaphase chromosomes to come much closer to the cortex of PP1-87B or Sds22 RNAi cells than they did in the control (Extended Data Fig. 5a, b; a similar phenotype was observed in cultured cells depleted for Sds22
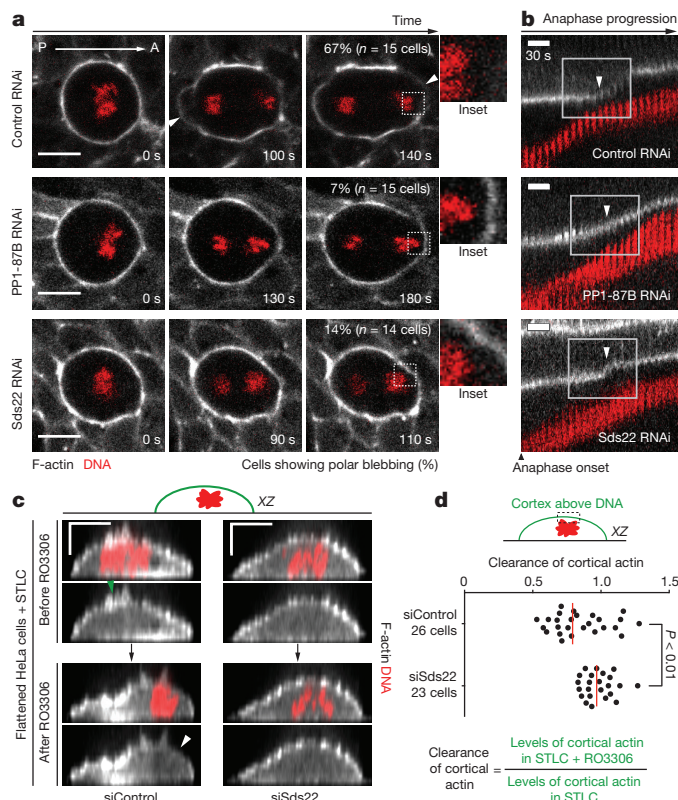
Figure 3 | PP1-87B and Sds22 phosphatase subunits are required for actin clearance and polar relaxation at mid anaphase. a, b, Polar actin clearance was analysed in control, PP1-87B RNAi and Sds22 RNAi SOP cells undergoing anaphase (images are representative of 15, 15 and 14 cells, respectively). Arrowheads in a indicate polar relaxation. Anaphase onset = 0 s. Scale bar, 5 μm. b, Kymographs for anterior pole of cells shown in a. Arrowheads indicate elongation onset. Grey box highlights polar relaxation (or impairment thereof). No significant clearance of polar actin was observed in PP1-87B-depleted cells (P = 0.099, 16 cells) or Sds22-depleted cells (P = 0.141, 10 cells) (as measured in Fig. 2e). c, d, XZ cross-sections of S-trityl-L-cysteine (STLC)-treated control or Sds22 RNAi HeLa cells (representative of 26 and 23 cells, respectively), expressing Rap1* before and after RO3306 addition, together with quantification of actin clearance from the cortex above the DNA. In c, the green arrowhead points to actin-enriched cortex and the white arrowhead points to actin-devoid cortex; scale bars, 10 μm. Lifeact–GFP labels F-actin in a–c. Median is shown in scatter plot (d). Significance was assessed using a two-tailed unpaired t-test.

Figure 4 | Local recruitment of Sds22 induces cortical relaxation. a–c, Cortical relaxation in Sds22–GFP-expressing fly cells in control and KNL1-depleted backgrounds. a, Representative images, together with b, their outlines over time, and c, graph of cell elongation showing mean and s.d. In a, inset regions are denoted by green arrowheads; black arrowhead points to polar blebbing. Asterisks mark DNA. Anaphase onset = 0 s. d, Representative images of HEK293T cells expressing CRY2–mCherry (31 cells), CRY–mCherry–Sds22 (39 cells) or CRY2–mCherry/Sds22–GFP (25 cells) following localized illumination with blue light (blue box). e, Model showing three anaphase signals that have been proposed to induce polarization of the actomyosin cortex: (1) Kinetochore-localized PP1–Sds22 inactivates moesin and triggers polar relaxation. (2) The spindle midzone promotes equatorial accumulation of myosin II and furrowing. (3) Chromatin-derived RanGTP is proposed to induce cortical repolarization and polar expansion[26]. Scale bars, 5 μm. Graphs show mean with s.d. Significance was assessed using a two-tailed unpaired t-test.

chromosome arms at the metaphase–anaphase transition[22]. However, Sds22–GFP was found concentrated at kinetochores (Extended Data Fig. 10a) during both metaphase and anaphase (Fig. 4a and Extended Data Fig. 10a–c). Moreover, the approach of kinetochore-bound Sds22–GFP to the cortex in mid anaphase was associated with local blebbing (Fig. 4a, b and Extended Data Fig. 10b, c). To determine whether this mid-anaphase polar relaxation depends on the observed accumulation of PP1–Sds22 at kinetochores, we followed chromosome segregation in KNL1 RNAi cells, in which the kinetochore recruitment of PP1–Sds22 is compromised[23]. The resulting drop in kinetochore localization of Sds22 (as measured by Sds22–GFP, Fig. 4a and Extended Data Fig. 10d, e) was associated with the loss of polar blebbing and with reduced anaphase cell elongation (Fig. 4b, c).

Finally, to test whether a spatially defined PP1 phosphatase signal is sufficient to trigger local relaxation of the cortex, as predicted by this model, we used optogenetics[24] to artificially target the subunit Sds22 to the plasma membrane in cells arrested in mitosis. By fusing cryptochrome 2 (CRY2) to Sds22 (CRY2–mCherry–Sds22), we were able to use light to drive the phosphatase to the membrane in cells co-expressing membrane-tethered CIBN[24] (a truncated version of the Arabidopsis cryptochrome-interacting basic helix-loop-helix protein) (see Methods and Extended Data Fig. 10f). In approximately half of the cells analysed, the local recruitment of Sds22–mCherry was associated with bleb formation (Fig. 4d). Thus, PP1–Sds22 is both required for anaphase cortical relaxation and is sufficient to drive this process when mislocalized to the cortex during metaphase.

(Extended Data Fig. 6b, c)). These defects were recapitulated by the expression of a constitutively active, phospho-mimetic form of moesin (Extended Data Fig. 7a–f), indicating that p-moesin is a key target of the phosphatase holo-enzyme. Moreover, PP1-87B complexes immunoprecipitated from fly cells efficiently dephosphorylated activated moesin in a manner that depended in part on Sds22 (Extended Data Fig. 7g, h). The block in polar relaxation induced by PP1–Sds22 silencing or the expression of constitutively active moesin led to a near-identical series of extreme contortions as cells entered telophase (Extended Data Fig. 8a–d and Supplementary Videos 4–7). However, cells nearly always completed a successful cytokinesis. This we attribute to the parallel recruitment of myosin II to the midzone, which is robust and functionally independent of PP1–Sds22 (ref. 18) (Extended Data Fig. 9a–c).

To determine the likely localization of the relevant PP1–Sds22 signal, we used an Sds22–green fluorescent protein (GFP) fusion to study the dynamic localization of PP1–Sds22 during mitotic progression. Several PP1 complexes are found at kinetochores, where they counteract Aurora B activity and silence the mitotic checkpoint[19–21]. Some, like Repo-Man–PP1, have been reported to move from the centromeres to
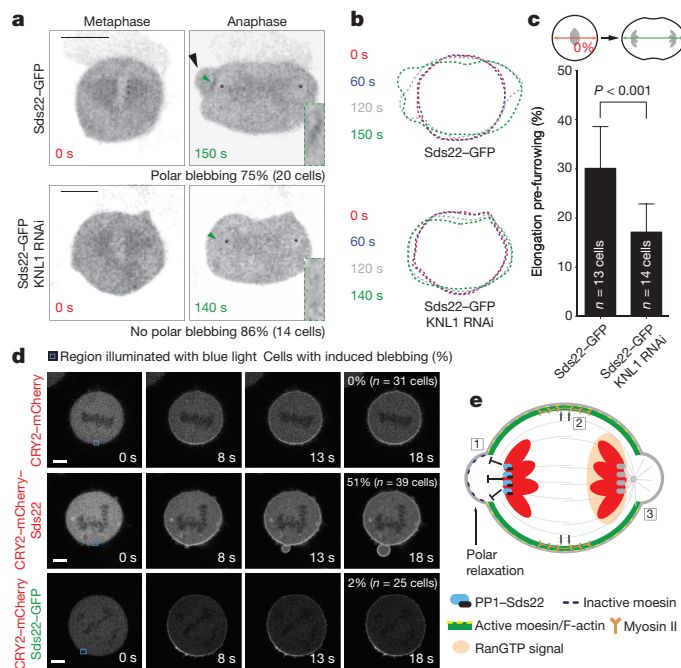
In summary, although work over many decades has focused on crosstalk between the central spindle and the overlying actomyosin cortex in positioning the cleavage furrow[1,9,18], this study reveals an additional, conserved kinetochore-based PP1–Sds22 signal that helps to couple chromosome segregation to anaphase cell elongation. Both signals are likely to act in parallel to generate the gradient in cortical actomyosin contractility required for cytokinesis[25] (Fig. 4e). The localization of PP1–Sds22 to kinetochores is striking, as related PP1 phosphatase complexes act at the same site to help control the decision to exit mitosis and to ensure the stable attachment of chromosomes to kinetochore microtubules[19–21]. As is the case for the PP1-mediated stabilization of microtubule–kinetochore attachments[20], the ability of kinetochore-localized PP1–Sds22 to act on the cortex is likely to depend on the phosphatase complex counteracting the activity of Aurora kinases. Unlike other phosphatase subunits[22], however, Sds22–GFP remains on kinetochores throughout anaphase. This places the PP1–Sds22 phosphatase complex in a perfect position to mediate communication between segregating chromosomes and the anaphase cortex, at least, in part, through the local dephosphorylation of ERM proteins at cell poles. Although recent studies have suggested a similar role for chromatin-bound Ran–GTP in cortical polarization during both meiotic and mitotic divisions[26,27] (Fig. 4e), the mechanistic link between Ran and the actomyosin cortex is not clear. Moreover, a function for Ran in this process is hard to reconcile with the enhancement of the polar relaxation signal seen following mitotic exit (Extended Data Fig. 4a, b). Therefore more work is required to determine how Ran, PP1–Sds22 and other signals emanating from the anaphase spindle function together to break cortical symmetry at mitotic exit, and how this initial difference between cortical contractility at cell poles and the midzone is amplified to drive furrow formation and cytokinesis, during both symmetric and asymmetric animal cell divisions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Green, R. A., Paluch, E. & Oegema, K. Cytokinesis in animal cells. *Annu. Rev. Cell Dev. Biol.* **28,** 29–58 (2012).
2. Kunda, P. *et al.* PP1-mediated moesin dephosphorylation couples polar relaxation to mitotic exit. *Curr. Biol.* **22,** 231–236 (2012).
3. Roubinet, C. *et al.* Molecular networks linked by Moesin drive remodeling of the cell cortex during mitosis. *J. Cell Biol.* **195,** 99–112 (2011).
4. Matthews, H. K. *et al.* Changes in Ect2 localization couple actomyosin-dependent cell shape changes to mitotic progression. *Dev. Cell* **23,** 371–383 (2012).
5. Sedzinski, J. *et al.* Polar actomyosin contractility destabilizes the position of the cytokinetic furrow. *Nature* **476,** 462–466 (2011).
6. Roegiers, F., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. Two types of asymmetric divisions in the *Drosophila* sensory organ precursor cell lineage. *Nature Cell Biol.* **3,** 58–67 (2001).
7. Georgiou, M. & Baum, B. Polarity proteins and Rho GTPases cooperate to spatially organise epithelial actin-based protrusions. *J. Cell Sci.* **123,** 1089–1098 (2010).
8. Hickson, G. R., Echard, A. & O'Farrell, P. H. Rho-kinase controls cell shape changes during cytokinesis. *Curr. Biol.* **16,** 359–370 (2006).
9. Mishima, M., Kaitna, S. & Glotzer, M. Central spindle assembly and cytokinesis require a kinesin-like protein/RhoGAP complex with microtubule bundling activity. *Dev. Cell* **2,** 41–54 (2002).
10. Lekomtsev, S. *et al.* Centralspindlin links the mitotic spindle to the plasma membrane during cytokinesis. *Nature* **492,** 276–279 (2012).
11. Tse, Y. C., Piekny, A. & Glotzer, M. Anillin promotes astral microtubule-directed cortical myosin polarization. *Mol. Biol. Cell* **22,** 3165–3175 (2011).
12. Fededa, J. P. & Gerlich, D. W. Molecular control of animal cell cytokinesis. *Nature Cell Biol.* **14,** 440–447 (2012).
13. Murthy, K. & Wadsworth, P. Dual role for microtubules in regulating cortical contractility during cytokinesis. *J. Cell Sci.* **121,** 2350–2359 (2008).
14. Werner, M., Munro, E. & Glotzer, M. Astral signals spatially bias cortical myosin recruitment to break symmetry and promote cytokinesis. *Curr. Biol.* **17,** 1286–1297 (2007).
15. Blachon, S. *et al.* *Drosophila asterless* and vertebrate Cep152 Are orthologs essential for centriole duplication. *Genetics* **180,** 2081–2094 (2008).
16. Hu, C. K., Coughlin, M., Field, C. M. & Mitchison, T. J. Cell polarization during monopolar cytokinesis. *J. Cell Biol.* **181,** 195–202 (2008).
17. Lancaster, O. M. *et al.* Mitotic rounding alters cell geometry to ensure efficient bipolar spindle formation. *Dev. Cell* **25,** 270–283 (2013).
18. Zhang, D. & Nicklas, R. B. 'Anaphase' and cytokinesis in the absence of chromosomes. *Nature* **382,** 466–468 (1996).
19. Lesage, B., Qian, J. & Bollen, M. Spindle checkpoint silencing: PP1 tips the balance. *Curr. Biol.* **21,** R898–R903 (2011).
20. Posch, M. *et al.* Sds22 regulates aurora B activity and microtubule-kinetochore interactions at mitosis. *J. Cell Biol.* **191,** 61–74 (2010).
21. Wurzenberger, C. *et al.* Sds22 and Repo-Man stabilize chromosome segregation by counteracting Aurora B on anaphase kinetochores. *J. Cell Biol.* **198,** 173–183 (2012).
22. Qian, J., Lesage, B., Beullens, M., Van Eynde, A. & Bollen, M. PP1/Repo-man dephosphorylates mitotic histone H3 at T3 and regulates chromosomal aurora B targeting. *Curr. Biol.* **21,** 766–773 (2011).
23. Liu, D. *et al.* Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase. *J. Cell Biol.* **188,** 809–820 (2010).
24. Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nature Methods* **7,** 973–975 (2010).
25. Turlier, H., Audoly, B., Prost, J. & Joanny, J. F. Furrow constriction in animal cell cytokinesis. *Biophys. J.* **106,** 114–123 (2014).
26. Kiyomitsu, T. & Cheeseman, I. M. Cortical dynein and asymmetric membrane elongation coordinately position the spindle in anaphase. *Cell* **154,** 391–402 (2013).
27. Dehapiot, B. & Halet, G. Ran GTPase promotes oocyte polarization by regulating ERM (Ezrin/Radixin/Moesin) inactivation. *Cell Cycle* **12,** 1672–1678 (2013).

METHODS

***Drosophila* strains and husbandry.** The following fly transgenes/alleles were used: *neur*-RFP (from Y. Bellaiche), UAS-LifeAct::GFP (from F. Schnorrer), *pnr*-GAL4 (BL3039), EM462-GAL4 (from G. Morata), *tub*-GAL80[ts] (BL7018 and BL7108), *neur*-GMA (GFP fused to the F-actin-binding domain of the fly moesin[28], expressed under the *neuralized* promoter), Sqh::mCherry (from E. Wieschaus), *ubi*-RFP::Centrosomin (from J. Raff), Asl[mecD] (from J. Raff), gSpc25::mRFP (from C. Lehner), UAS-Sds22::GFP (from B. Thompson) and UAS-Moesin::GFP and UAS-MoesinT559D::GFP (from F. Payre). RNA interference (RNAi) lines were used to silence the expression of the following genes: *RacGAP1/tumbleweed* (DRSC, HMS01417), *Sds22* (VDRC, TID42051), *PP1-87B* (VDRC, TID35024) and *KNL1/Spc105R* (DRSC, HMS01752). To drive RNAi expression, we first generated a fly strain encoding *neur*-RFP (which labels the chromatin of sensory organ cells at all stages of the cell cycle), *pnr*-GAL4 (*pannier* is expressed in the central region of the notum[29]) and UAS-LifeAct::GFP (to visualize actin filaments). Females of this line were then crossed to UAS-hairpin RNAi males. To overcome the lethality associated with expression of the inducible hairpin targeting either PP1-87B or RacGAP1, we used a GAL80 temperature-sensitive background[30]. All flies were grown at 18 °C and shifted to 29 °C (to inactivate GAL80[ts] allowing GAL4-dependent gene transcription) at the beginning of pupariation. Expression of the inducible hairpin targeting KNL1/Spc105R was accomplished using the EM462-GAL4 driver.

Cell lines. Well-established cell lines were used, all of which have been used in previously published studies from the Baum and Hickson laboratories and their collaborators. Human cell lines were subjected to frequent mycoplasma tests and tested negative relative to a positive control.

Plasmids for expression in human cells. To create CRY2–mCherry–Sds22, the Sds22 coding sequence (NM_002712) was inserted into CRY2–mCherry by using sequence- and ligation-independent cloning as previously described[31]. Briefly, a CRY2–mCherry vector was linearized by inverse PCR reaction. Sds22 was amplified by PCR with primers containing a 20-bp extension homologous to each end of the linearized vector. Vector and insert were mixed and incubated with T4 polymerase, followed by transformation in TOP10-competent cells (Invitrogen).

siRNA transfection. The following siRNA duplexes were used at a final concentration of 50 nM: AllStars control (Qiagen, 1027280), RacGAP1 (Invitrogen Stealth HSS120934) (GCCAAGAACUGAGACAGACAGUGUG) and Sds22 (Qiagen Hs_PPP1R7_5) (CCAGATCAAGAAGATTGAGAA). Lipofectamine RNAiMAX (Invitrogen) was used for siRNA transfection. Cells were analysed 31 h after transfection with siRNA. For the experiments of cell flattening (in Fig. 3c, d, Extended Data Fig. 4a, b and Extended Data Fig. 6d), cells were transfected with Rap1* before siRNA treatment, as previously described[17].

Live imaging of *Drosophila* SOP cells. For time-lapse imaging, fly pupae were first attached to a slide with double-sided tape on their ventral side, as previously described[7]. A small window was then cut in the pupal case on the dorsal side to expose the notum. A coverslip, together with a drop of halocarbon oil, was then placed over the whole fly, supported by coverslips glued onto either end of the slide to allow imaging on confocal microscopes. Time-lapse movies were acquired using a Leica SPE confocal microscope with a 40× lens (NA 1.3). Nota were imaged at 14–16 h after puparium formation. In all cases, male and female flies were treated equivalently without randomization; animal numbers used were determined by experimental constraints.

***Drosophila* S2 cell experiments.** Schneider's S2 cells (from the O'Farrell laboratory, University of California, San Francisco), grown in Schneider's medium supplemented with 10% fetal calf serum (Life Technologies), were seeded in 8-well Labtek chambered coverglass dishes (Thermo Fisher Scientific) and treated with 25 μM colchicine (Sigma) for 30 min before the start of imaging or fixation. Mitotic exit was induced by prior incubation for 48 h with Mad2 double-stranded RNAs (dsRNAs) (as previously described[32]) or by 10 μM RO3306 for 5–10 min. Cells were fixed using 4% formaldehyde in PBS for 10 min before permeabilization and blocking with PBS containing 0.1% triton X-100 and 5% normal goat serum (Jackson ImmunoResearch). Cells were immunostained using antibodies against p-ERM (Cell Signaling, 3141S, 1/200) and anillin (1/1,000, a gift from C. Field). DNA was stained using Hoechst 33258 (1/1,000) and F-actin was stained with Alexa546-conjugated phalloidin (Molecular Probes). Live-cell imaging was performed using cells stably expressing Histone-H2B–GFP and anillin–mCherry constructs under the control of the constitutive *act5C* promoter. Imaging was performed at room temperature using an Ultraview Vox spinning disc confocal system (PerkinElmer), employing a CSU-X1 scanning unit (Yokogawa) and an Orca-R2 charge-coupled device camera (Hamamatsu) fitted to a Leica DMI6000B inverted microscope equipped with a motorized piezo-electric stage (Applied Scientific Instrumentation). Image acquisition was performed using Volocity 6 (Improvision/Perkin Elmer) and a Plan Apo 63× oil-immersion objective (NA 1.4) with camera binning set to 2 × 2.

Live imaging of human cells. HEK293T cells (gift from the Marsh laboratory, University College London) were plated on glass-bottomed dishes (MatTek). HeLa cells stably expressing LifeAct–GFP/Histone-H2B-mRFP (as previously published[4,17]) were plated on chambered coverslips (Labtek; Thermo Fisher Scientific) coated with 10 μg ml⁻¹ fibronectin (Sigma). Before recording of live cells, the medium was changed to a phenol-red-free CO₂-independent medium, L-15 (Leibovitz). For Fig. 1c, Fig. 3c, d, Extended Data Fig. 4a, b and Extended Data Fig. 6, frames were acquired at 37 °C using an UltraView Vox (Perkin Elmer) spinning disc confocal microscope with a 60× oil objective (NA 1.4) and images acquired using a Hamamatsu C9100-13 EMCCD camera and Volocity software (Perkin Elmer). Images in Fig. 1d and Extended Data Fig. 2b were acquired at 37 °C using a Zeiss Axio Observer Z1 microscope controlled by Volocity 6.3 software (Perkin Elmer) and equipped with an Orca 03GO1 camera (Hamamatsu) and a 40× oil DICII objective (NA 1.3). For the optogenetic experiments (Fig. 4d), HEK293T cells were co-transfected with CIBN–GFP–CAAX, CRY2–mCherry, CRY2–mCherry–Sds22 and Sds22–GFP (plasmid from J. Swedlow) using FuGENE HD (Promega). By 22 h after transfection, cells were treated with 50 ng ml⁻¹ nocodazole for 8 h and imaged after nocodazole wash-out using a Leica TCS SP5 confocal microscope and 60× oil objective (NA 1.4). Ten consecutive pulses (800 ms each) of 488-nm light at 15% of laser output were used to induce interaction between CIBN and CRY2. A 561-nm laser was used to image CRY2–mCherry or CRY2–mCherry–Sds22.

Immunofluorescence microscopy in *Drosophila*. Nota from pupae 14–16 h after puparium formation were dissected in PBS at room temperature. Tissues were promptly fixed in 10% trichloroacetic acid or 4% formaldehyde for 20 min at room temperature, before being permeabilized in PBS containing 0.1% triton X-100. Subsequently, nota were incubated in a blocking solution composed of 5% bovine serum albumin (Sigma) and 3% fetal bovine serum (PAA) (in PBS). Once immunostained and mounted, tissues were imaged using a Leica TCS SP5 confocal microscope and 60× oil objective (NA 1.4).

Immunofluorescence microscopy in human cells. For the data shown in Extended Data Fig. 3f, hTERT RPE-1 cells (used previously[4]) were arrested in prometaphase by using the Eg5 kinesin inhibitor STLC (5 μM, Sigma) for 8 h. Cells were treated with a low dose of nocodazole (20 nM, Sigma) for 3 h to specifically disrupt astral microtubules[33]. To induce monopolar cytokinesis, the CDK1 inhibitor RO3306 (10 μM, Sigma) was added to STLC-treated cells for 10 min. Cells were then fixed on ice in 10% trichloroacetic acid for 15 min before being processed for immunofluorescence microscopy as described[34]. Images in Extended Data Fig. 3f were acquired on a Leica SPE confocal microscope with a 63× lens (NA 1.3).

Immunoprecipitation assays. For the data shown in Extended Data Fig. 7g, dsRNAs targeting LacZ and Sds22 were generated using T7-tagged oligonucleotides and a T7 dsRNA kit (Ambion) as previously described[35]. S2 cells were transfected with dsRNA following transfection with PP1-87B–GFP on the next day. Cells were lysed in buffer containing 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5% triton X-100 and a protease inhibitor cocktail (Roche) 72 h after dsRNA transfection. PP1-87B–GFP was collected with GFP–Trap-A (Chromotek). For the substrate, S2 cells were transfected with Flag–moesin. By 48 h after transfection, cells were treated with 50 nM calyculin A for 25 min and lysed in buffer containing 10 mM Tris HCl pH 7.5, 150 mM NaCl, 1% triton X-100, 0.1% SDS, 1% sodium deoxycholate, a protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktails 2 and 3 (Sigma). Flag–moesin was collected with anti-Flag M2 beads (Sigma) and eluted with buffer containing 50 mM Tris HCl pH 7.5, 30 mM NaCl and 150 ng μl⁻¹ 3xFlag peptide. Eluted Flag–moesin was added to the beads bound to PP1-87B–GFP, incubated at 30 °C for the indicated time and analysed by immunoblotting.

Antibodies and dyes. The following antibodies and dyes were used for immunofluorescence microscopy: rabbit anti-p-ERM (Cell Signaling, 3141S, 1/100 in fly tissues, 1/200 in human cells), rabbit anti-pMRLC2 (S19) (Cell Signaling, 3671S, 1/100), mouse anti-α-tubulin (Abcam, DM1A, 1/100 in fly tissues, 1/200 in human cells), guinea pig anti-centrosomin (1/1,000, a gift from F. Pichaud), 4′,6-diamidino-2-phenylindole (DAPI) (Molecular Probes, 1 μg ml⁻¹), Phalloidin-TRITC (Sigma-Aldrich). Secondary antibodies from Molecular Probes were conjugated with Alexa488, 546 and 647.

The following antibodies were used for western blotting: rabbit anti-p-ERM (Cell Signaling, 3141S, 1/100), rabbit anti-moesin (1/25,000, a gift from D. Kiehart), mouse anti-GFP (Roche, No. 11814460001, 1/500), rabbit anti-Sds22 (Santa Cruz, E-20, 1/100), mouse anti-γ-adaptin (BD Transduction Laboratories, No. 610386, 1/1,000) and mouse anti-RacGAP1 (Abnova, M01, 1/50,000).

Quantification of protein levels at the cortex. FIJI software (http://fiji.sc/Fiji) was used to quantify the actin levels at the cell pole during anaphase (as seen in Fig. 1f, Fig. 2e and Extended Data Fig. 1d). First, a kymograph of the radial cross-section was assembled for each cell (as seen in Fig. 2a, b, yellow box denotes the

cross-section). Then, the mean actin intensity was systematically measured over time by placing a 9-pixel square region onto the polar cortex at every time point along the kymographic profile. Values were then normalized against the actin levels in metaphase. To generate the bar graphs of the actin levels, the values obtained for each cell represent the average of five values measured during a 50-s period either before or after the onset of cell elongation.

To quantify the poles/equator (or proximal/distal) ratios of protein levels seen in Fig. 1h, Extended Data Fig. 1e and Extended Data Fig. 3h, i, protein levels were measured by placing a small rectangle onto the polar or equatorial cortex of anaphase cells and retrieving the mean grey value (using FIJI). In Fig. 3d, Extended Data Fig. 2c, d and Extended Data Fig. 4a, levels of cortical protein (F-actin or myosin II) were normalized against the cytoplasmic signal. To quantify the levels of F-actin seen in Extended Data Fig. 4b and Extended Data Fig. 6d, the levels of cortical protein under the chromatin mass were normalized against the levels of cortical signal of the whole cell area (bottom cross-section).

Assembly of plots: actin levels vs time and DNA-to-cortex distance vs time. FIJI software (http://fiji.sc/Fiji) was used to generate kymographs of the cortical perimeter over time (as seen in Fig. 2g, top left). A line scan was drawn over the half-perimeter of each Lifeact–GFP-labelled SOP cell during early anaphase (nine time points, 10-s resolution). These lines were then straightened and combined in the form of a kymograph (Fig. 2g, top left). These kymographs were saved in a text image format and subsequently binned into 9×9 grids of mean grey values (not shown). These 9×9 grids of actin intensity values were then 'overlaid' to obtain an average plot of actin levels vs time for all cells analysed (seen in Fig. 2g, bottom left). In parallel to this, we measured the DNA–cortex distances, d1–d9, (as seen in Fig. 2f) in each cell during early anaphase, obtaining a 9×9 grid of distance vs time as seen in Fig. 2g, top right). Similarly to the actin plots, the distance vs time grids of all ten cells analysed were 'overlaid' to obtain an average plot of distance vs time (seen in Fig. 2g, bottom right). Matlab (MathWorks) was used to present the aforementioned plots in the form of colour maps.

Graphs and statistical analysis. Various types of graphs are shown: bar graphs of mean with s.d. (Figs. 1b, d, f; 2e; 4c; Extended Data Figs. 1c, d; 2c, d; 3c, k; 5h, k; 7f; 9c; 10e), scatter plots with median (Fig. 3d; Extended Data Figs. 1b; 3h, i; 6c, d) and box-and-whisker plots with 10–90 percentiles (Fig. 1h; Extended Data Figs. 1e; 4a, b; 5b, d, e; 7c; 8c, d). Welch's unequal variances t-test was used to calculate statistical significance in all analyses. $P < 0.05$: significant; $P < 0.01$: very significant; $P < 0.001$: extremely significant. Linear regression analysis was carried out using Graphpad PRISM. Note that independent data points represent biological replicates: that is, parallel experiments carried out on genetically identical cells with identical reagents. Figure legends indicate when the same primary data (images/movies) were subjected to different types of analysis in parallel.

28. Edwards, K. A., Demsky, M., Montague, R. A., Weymouth, N. & Kiehart, D. P. GFP-moesin illuminates actin cytoskeleton dynamics in living tissue and demonstrates cell shape changes during morphogenesis in Drosophila. *Dev. Biol.* **191,** 103–117 (1997).
29. Mummery-Widmer, J. L. *et al.* Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. *Nature* **458,** 987–992 (2009).
30. Matsumoto, K., Toh-e, A. & Oshima, Y. Genetic control of galactokinase synthesis in *Saccharomyces cerevisiae*: evidence for constitutive expression of the positive regulatory gene *gal4*. *J. Bacteriol.* **134,** 446–457 (1978).
31. Jeong, J. Y. *et al.* One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl. Environ. Microbiol.* **78,** 5440–5443 (2012).
32. Hickson, G. R. & O'Farrell, P. H. Rho-dependent control of anillin behavior during cytokinesis. *J. Cell Biol.* **180,** 285–294 (2008).
33. Théry, M. *et al.* The extracellular matrix guides the orientation of the cell division axis. *Nature Cell Biol.* **7,** 947–953 (2005).
34. Lénárt, P. *et al.* The small-molecule inhibitor BI 2536 reveals novel insights into mitotic roles of polo-like kinase 1. *Curr. Biol.* **17,** 304–315 (2007).
35. Liu, T., Sims, D. & Baum, B. Parallel RNAi screens across different cell lines identify generic and cell type-specific regulators of actin organization and cell morphology. *Genome Biol.* **10,** R26 (2009).
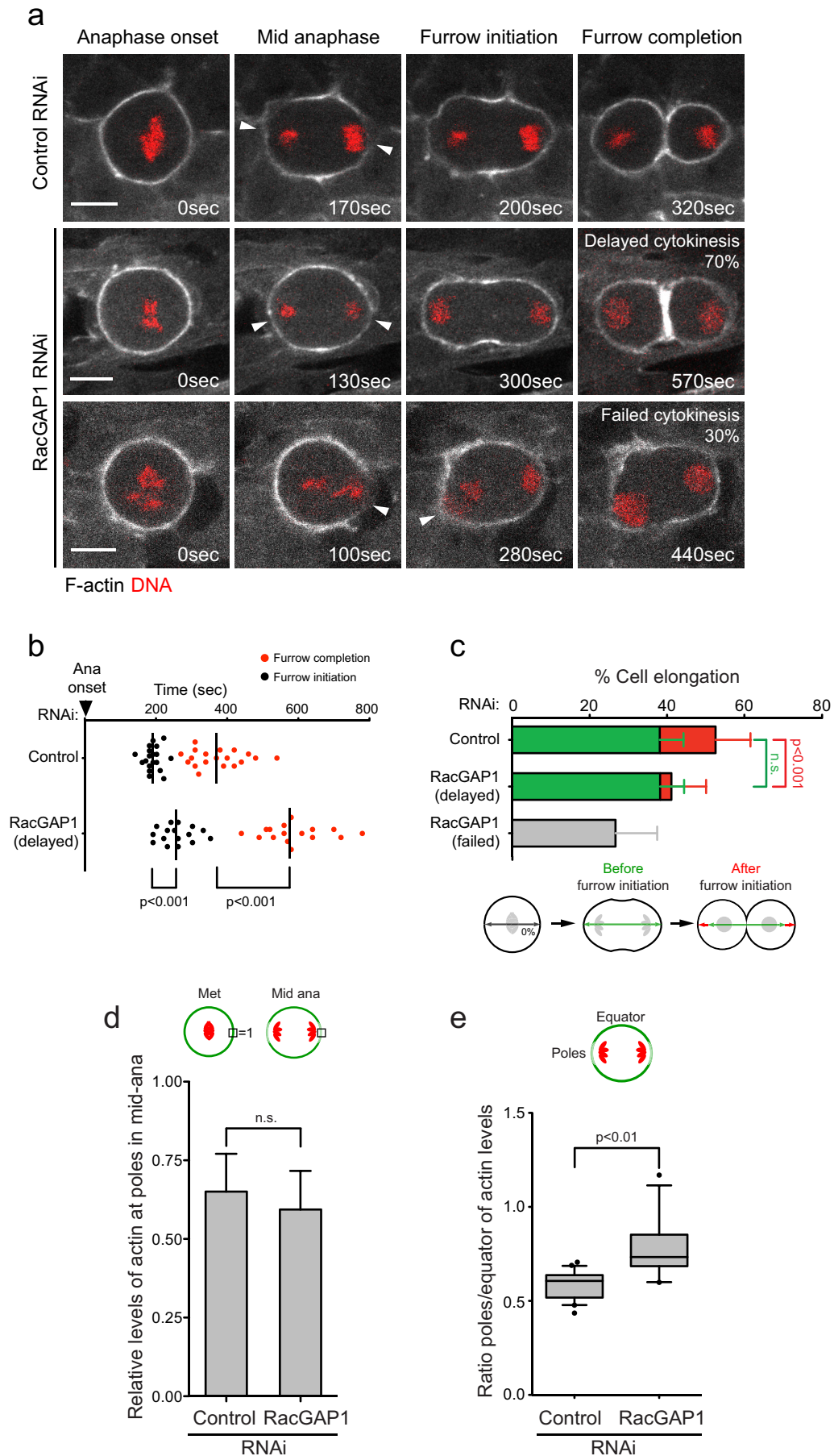
Extended Data Figure 1 | Depletion of RacGAP1 in SOP cells affects neither polar relaxation nor anaphase cell elongation. a–c, Time-lapse imaging of SOP cells in control and RacGAP1-depleted backgrounds was carried out to analyse the relative timing of polar relaxation and anaphase elongation. We analysed cells from five control animals and four RacGAP1 RNAi animals. Representative images are shown in **a** (scale bars, 5 μm). White arrowheads indicate actin clearance at the poles. Lifeact–GFP was used to label F-actin. For control and RacGAP1 RNAi cells that formed a furrow ($n = 19$ and $n = 16$, respectively), the period between anaphase onset and furrow initiation or furrow completion is plotted in **b** (black lines denote the median). Cell length is plotted in **c** for the control group (18 cells) and for RacGAP1 RNAi cells that were delayed in furrow formation (15 cells) or that failed to form a furrow (5 cells). **d**, Relative levels of polar actin were compared across movies of 12 control RNAi cells from four animals (same as seen in Fig. 2e) and from 12 RacGAP1 RNAi cells from three animals. Ana, anaphase; Met, metaphase. **e**, Graph shows the ratio of levels of cortical actin at poles versus the equator at mid anaphase for 20 control cells from five animals, and for 13 RacGAP1 RNAi cells from three animals. Data are shown as mean and s.d. in **c** and **d**, and as box-and-whisker plots with 10–90 percentiles in **e**. A two-tailed unpaired t-test was used to calculate statistical significance; $P > 0.05$ was deemed not significant (n.s.).
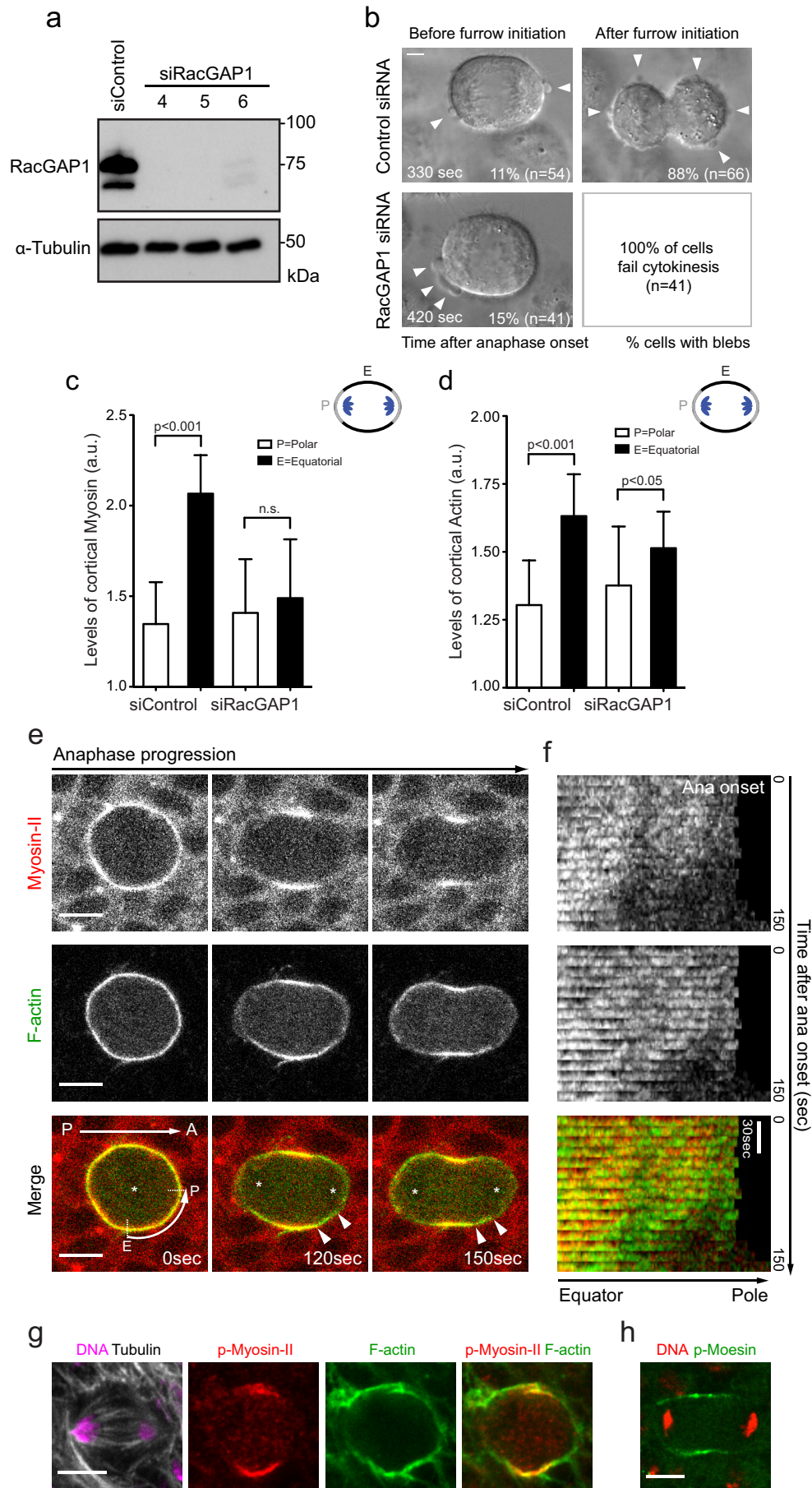
Extended Data Figure 2 | Depletion of RacGAP1 impairs myosin II equatorial accumulation and furrow ingression, but does not affect actin clearance from the poles at mid anaphase. **a**, Western blot showing RacGAP1 depletion in HeLa cells. **b**, Representative DIC stills from movies show HeLa cells at indicated times after the onset of anaphase. Images show control cells before and after furrow initiation and RacGAP1-depleted cells at mid anaphase, as all fail cytokinesis. Arrowheads point to blebs. **c**, **d**, Graphs show levels of cortical myosin II (**c**) and cortical actin (**d**) in the polar and equatorial regions of cells. Levels were measured in HeLa cells expressing myosin-II–GFP and Utrophin–Ruby treated with control siRNA (15 cells, 3 experiments) or siRacGAP1 siRNA (oligonucleotide no. 4, seen in **a**) (15 cells, 3 experiments). Data are shown as mean and s.d. **e**, **f**, Representative images and corresponding kymographs taken from 12 time-lapse movies of fly SOP cells, fluorescently labelled for both myosin II (Sqh–mCherry) and actin filaments (GMA), undergoing anaphase. Note that the same cells were used for the analysis in Extended Data Fig. 9a. Anaphase onset = 0 s. Asterisks mark the chromosomes. Kymographs of anaphase progression of the E–P perimeter section depicted in **e**. Note that actin and myosin II have distinct patterns of localization at the cortex during anaphase (also, see white arrowheads in **e**). **g**, Fly epithelial cell at mid anaphase immunostained for tubulin, DNA, p-myosin II and F-actin (phalloidin), representative of three cells. **h**, Fly epithelial cell at mid anaphase immunostained for p-moesin and DNA, representative of 15 cells. Scale bars in **b**, **e**, **g**, **h**, 5 μm. A two-tailed unpaired *t*-test was used to calculate statistical significance.

a

Anterior pole

Meta | Ana

Cnn F-actin DNA

Anaphase progression

30sec

t1 | t2 | t1 | t2

b

Control

Tubulin DNA Cnn

Asl^{mecD}

c

p<0.001

Cnn levels at centrosome (a.u.)

Control | Asl^{mecD}

d

P → A

0°
30°
30°
90°

SOP cells

Number of cells

Control

Number of cells

Asl^{mecD}

e

Early ana | Mid ana

Control

Asl^{mecD}

F-actin Cnn

Distance (a.u.) | Distance (a.u.)

Actin levels

f

STLC-treated RPE1 cells

+RO3306

+RO3306 | +RO3306 +low Noco

p-Moesin Tubulin DNA

g

F-actin | Anillin | Merge

Colchicine-treated Mad2 RNAi S2 cells

F-actin | p-Moesin | Merge

Colchicine + RO3306-treated S2 cells

j

DNA Anillin

0sec | 180sec | 300sec | 82%

0 Time (sec) 300

Colchicine-treated Mad2 RNAi S2 cells

DNA F-actin

0sec | 120sec | 240sec | 83%

0 Time (sec) 360

Colchicine-treated Mad2 RNAi S2 cells

h

Proximal | Distal

Ratio proximal/distal of F-actin levels

n=13

i

Proximal | Distal

Ratio proximal/distal of p-Moesin levels

n=13

k

Mitotic Exit

I.
II.
III.

○ Anillin ✳ DNA

% of Colchicine-treated S2 cells

I II III

Mad2 RNAi | RO3306

Extended Data Figure 3 | Actin clearance from the poles is independent of centrosomes and astral microtubules. a, SOP cell imaged at metaphase and anaphase (left) (representative of three cells imaged precisely in this way), together with a kymograph of the cross-section (yellow box). Cnn indicates centrosomin. Lifeact–GFP was used to label F-actin. **b, c**, Fly epithelial cells were fixed and immunostained for centrosomin, tubulin and DNA. Green arrowheads indicate the presence of centrosomes in control cells. Representative images are shown (**b**), together with a quantification of relative centrosomin levels at the centrosome (**c**) for 25 cells from three control animals and for 26 cells from three Asl^{mecD} animals. A two-tailed unpaired t-test indicated that there was a significant difference in centrosomal centrosomin levels in the two cases. **d**, Scheme of SOP cells dividing in different orientations. A–P axis = 0° (left). Rosette plots indicate spindle axis angle measured at the onset of anaphase for 34 control cells from three animals and for 23 Asl^{mecD} cells from three animals. **e**, Time-lapse stills of SOP cells expressing GMA to label F-actin taken at early and mid anaphase (ana) in control (representative of 12 cells) and Asl^{mecD} (representative of 16 cells) mutant backgrounds (as shown in Fig. 1e, f), together with plot profiles (right) denoting the relative actin levels across the cell. Asterisks mark the chromosomes. **f**, Images show representative STLC-treated RPE-1 cells treated with or without 20 nM nocodazole (Noco) and/or RO3306 (15 cells were analysed for each condition), fixed and stained for p-ERM proteins, DNA and tubulin. **g**, Images in top panel show representative Mad2-depleted S2 cells treated with 25 μM colchicine and stained for F-actin (phalloidin, red in merged image), anillin (green in merged image) and Hoechst 33258 (blue in merged image), from a population of 13 cells. Similarly, the bottom panel shows images of S2 cells (representative of 13 cells) treated with colchicine, and 20 μM RO3306 to induce forced exit from mitosis, and stained for F-actin (phalloidin, red in merged image) and p-moesin (p-ERM antibody, green in merged image) and Hoechst 33258 (blue in merged image). **h, i**, Ratio of proximal/distal levels of cortical F-actin (**h**) and p-moesin (**i**) (refers to **g**). Mean is labelled in red. **j**, S2 cells, expressing either H2B–GFP/anillin–Cherry or Lifeact–GFP/H2B–Cherry, were imaged during mitotic exit. Representative stills and the corresponding kymographs are shown in **j** (equivalent to phenotype I in **k**). Dashed lines were used to generate the kymographs. Top panel, $n = 68$ cells, three experiments; bottom panel, $n = 24$ cells, one experiment. **k**, Phenotypic quantification of anillin–Cherry-expressing S2 cells treated with colchicine and forced to exit mitosis through either Mad2 depletion (as depicted in **j**, top panel) or through treatment with RO3306. Bar graphs depict mean and s.d. Phenotype I: DNA and cortex are polarized. Phenotype II: neither DNA nor cortex is polarized. Phenotype III: DNA is polarized but cortex is not. Mad2 RNAi, $n = 68$ cells, three experiments. RO3306, $n = 121$ cells, two experiments. In **a, b, f, g** and **j**, scale bars, 5 μm.

a STLC

Actin levels above DNA

=1.27

=1.02

=0.89

=0.79

F-actin DNA | F-actin

p<0.05

Levels of cortical actin above DNA (a.u.)

n=17 cells

STLC | STLC + RO3306

b STLC

Actin levels below DNA

=0.98

=0.80

XY (bottom cross-section)

F-actin DNA | F-actin

p<0.01

Levels of cortical actin below DNA (a.u.)

n=17 cells

STLC | STLC + RO3306

c D1 D2 D3

e c1 c2 c3 c4 c5 c6 c7 c8 c9

d Cell A

Ana onset | Elongation onset | Furrow initiation

DNA-to-cortex (μm) | Cell center-to-cortex (μm) | DNA-to-cell center (μm)

D1 D2 D3

Time after anaphase onset(sec)

f Posterior pole (mirror image)

F-actin DNA

−80 −40 0

Time to expansion of anterior pole (sec)

g Anterior pole

F-actin DNA

−80 −40 0

Time to expansion of anterior pole (sec)

h Plot actin levels/time (Post pole)

c1 c2 c3 c4 c5 c6 c7 c8 c9

10 cells

Actin levels (a.u.)

Time to expansion of anterior pole (sec)

Timing of posterior pole expansion: 113±24sec after Ana onset (10cells)

i Plot actin levels/time (Ant pole)

c1 c2 c3 c4 c5 c6 c7 c8 d9

10 cells

Actin levels (a.u.)

Time to expansion of anterior pole (sec)

Timing of anterior pole expansion: 98±18sec after Ana onset (10cells)

Extended Data Figure 4 | DNA-induced clearance of cortical F-actin at anaphase. a, b, Data show representative stills and corresponding quantitative data extracted from movies of 17 STLC-treated HeLa cells (from three independent experiments) expressing LifeAct–GFP and H2B–Cherry forced to flatten through Rap1* expression before and after treatment with the CDK inhibitor RO3306. **a,** Images show *XZ* cross-sections of a representative flattened HeLa cell before and after treatment with RO3306. Levels of cortical F-actin above the chromatin were normalized against the cytoplasmic fluorescence signal (ratios are shown in green on right). Graph on right shows normalized levels of cortical actin overlying the DNA before and after treatment with RO3306 (at 6 min after drug addition) for all 17 cells. **b,** *XY* cross-sections of representative cell shown in **a** (left). Levels of cortical actin below the chromatin (see dotted region) were normalized against the fluorescence signal in the most basal confocal section (ratios shown in green on right). Graph on right shows normalized levels of basal cortical actin lying beneath the DNA for all 17 cells. **c, d,** Scheme and data to test the correlation between cell elongation and anaphase chromosome movements at the anterior pole of fly SOP cells. **c,** Scheme depicts distances D1, D2 and D3. **d,** Graph shows D1, D2 and D3 plotted for anterior pole during anaphase for representative SOP cell A (shown in Fig. 2a, b, 1 of 12 analysed). **e–i,** Experiments to test how cortical actin is cleared from the anterior and posterior cortex of ten SOP cells during chromosome segregation. **e,** Scheme of cortical regions c1–c9 (as seen in Fig. 2f). **f, g,** Stills of the posterior and anterior poles of a representative SOP cell imaged in early anaphase. Arrowheads in **f** and **g** point to poor and strong actin clearance, respectively. **h, i,** Average plot of cortical actin measured over time for the posterior pole (Post) and anterior (Ant) pole (as seen in Fig. 2g, bottom left). The F-actin threshold level was set to 3.0 to facilitate the comparison between anterior and posterior poles. These data show that clearance of actin on the anterior pole occurs before clearance on the posterior pole in SOP cells. Scale bars in **f** and **g**, 5 μm. Box-and-whisker plots show median together with 10–90 percentiles. A two-tailed unpaired *t*-test was used to calculate statistical significance.
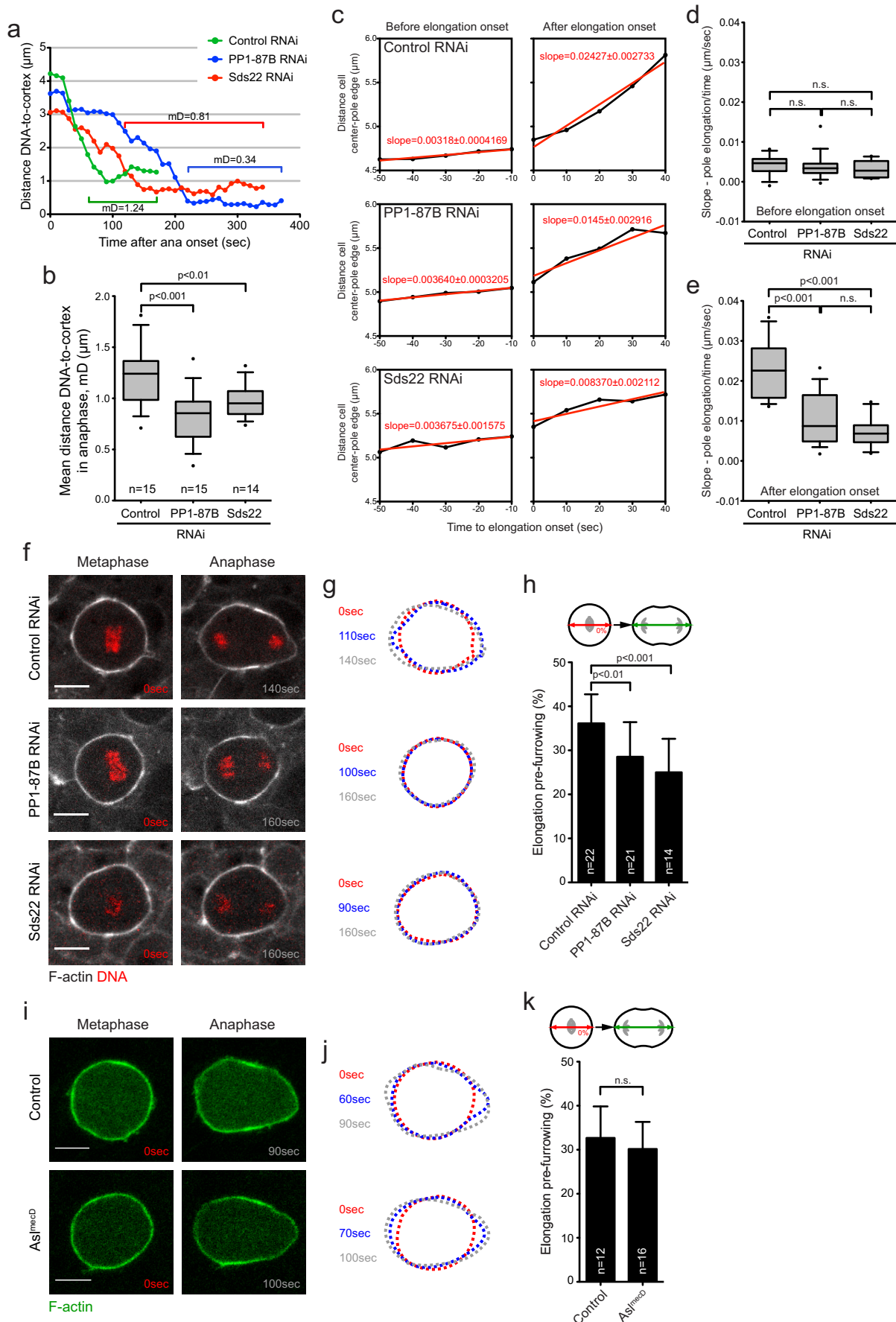
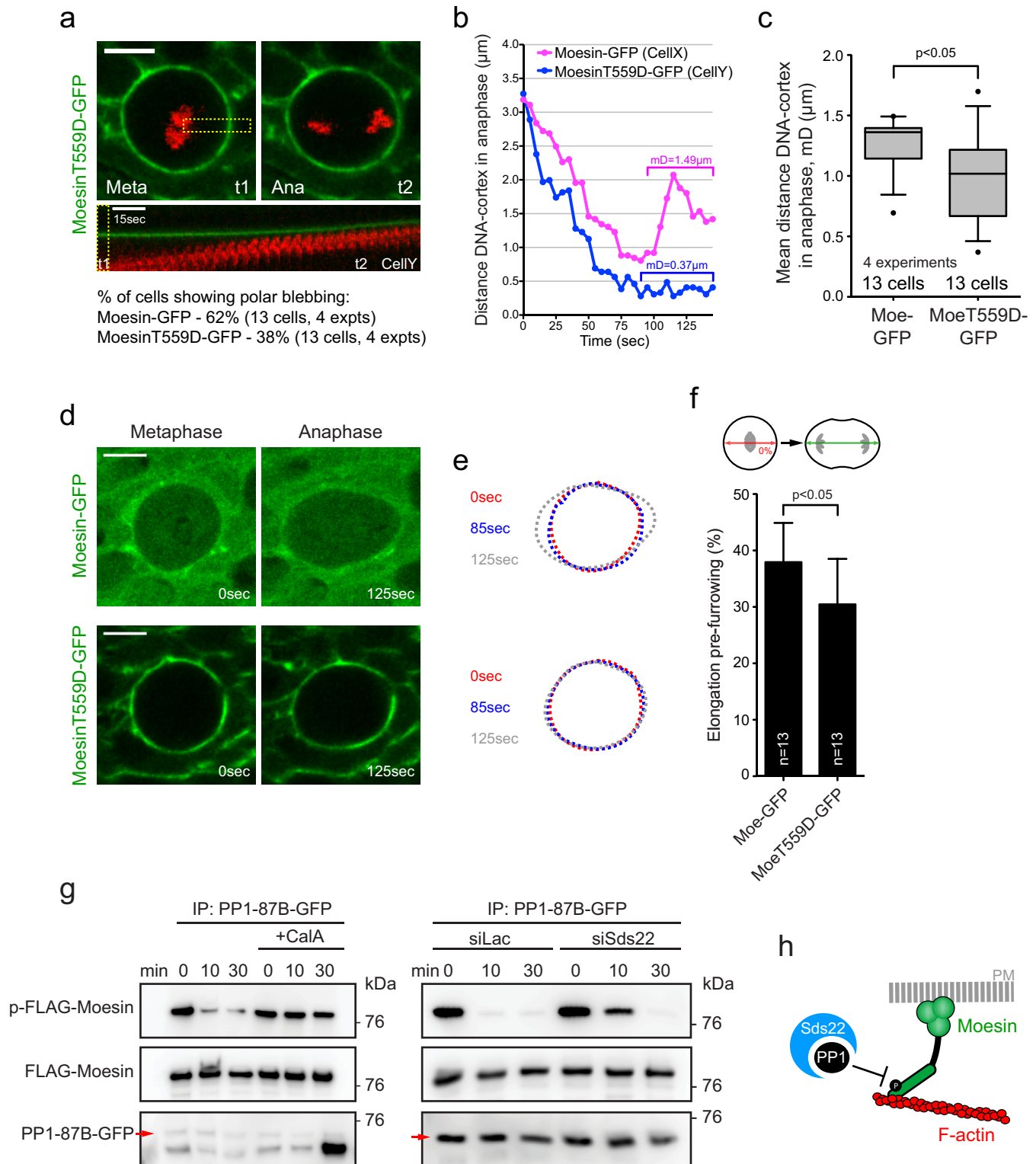**Extended Data Figure 5 | Depletion of PP1-87B or Sds22 impairs cell elongation in SOP cells. a–e**, The correlation between cell elongation and the approach of chromatin to the cortex was analysed in control (12 cells, four animals), PP1-87B RNAi (16 cells, four animals) and Sds22 (10 cells, three animals) RNAi cells. **a**, Plot of the DNA-to-cortex distance during anaphase in three representative SOP cells in control, PP1-87B RNAi and Sds22 RNAi backgrounds. Anterior pole depicted. mD, mean distance during anaphase. **b**, Box plot of mean DNA-to-cortex distance in mid anaphase. **c**, Graphs show distance from cell centre to pole plotted before and after elongation onset in representative cells for each of the three conditions (in black), together with a fitted linear regression (in red). **d, e**, Box plot to show the slopes of linear regression analysis (as in **c**) before the elongation onset and after the elongation onset for control, PP1-87B RNAi and Sds22 RNAi backgrounds. **f–h**, Pre-furrow anaphase elongation for control (22 cells from five animals), PP1-87B RNAi (21 cells from four animals) and Sds22 RNAi (14 cells from three animals) SOP cells expressing Lifeact–GFP. **f**, Representative images of cells.

g, Outlines of the boundary of cells shown in **f** at different times following the onset of anaphase. **h**, Graph shows pre-furrow anaphase cell elongation for cells in each background. These data show that PP1-87B or Sds22 depletion leads to a defect in anaphase elongation in SOP cells. **i–k**, Analysis shows anaphase elongation in control (12 cells from three animals) and Asl^{mecD} mutant (16 cells from three animals) cells. **i**, Images show F-actin in representative SOP cells expressing GMA in control and Asl^{mecD} mutant backgrounds. **j**, Outlines of boundary at different times following anaphase onset for representative cells shown in **i**. **k**, Plot of cell elongation in the backgrounds seen in **i**. These data show that anaphase cell elongation is not perturbed in the absence of centrosomes or astral microtubules. n, number of cells. Control, three animals; Asl^{mecD}, three animals. In **f** and **i**, scale bars, 5 μm. Box-and-whisker plots show median and 10–90 percentiles. Bar charts show mean and s.d. Significance was assessed using a two-tailed unpaired t-test. $P > 0.05$ was deemed not significant (n.s.).
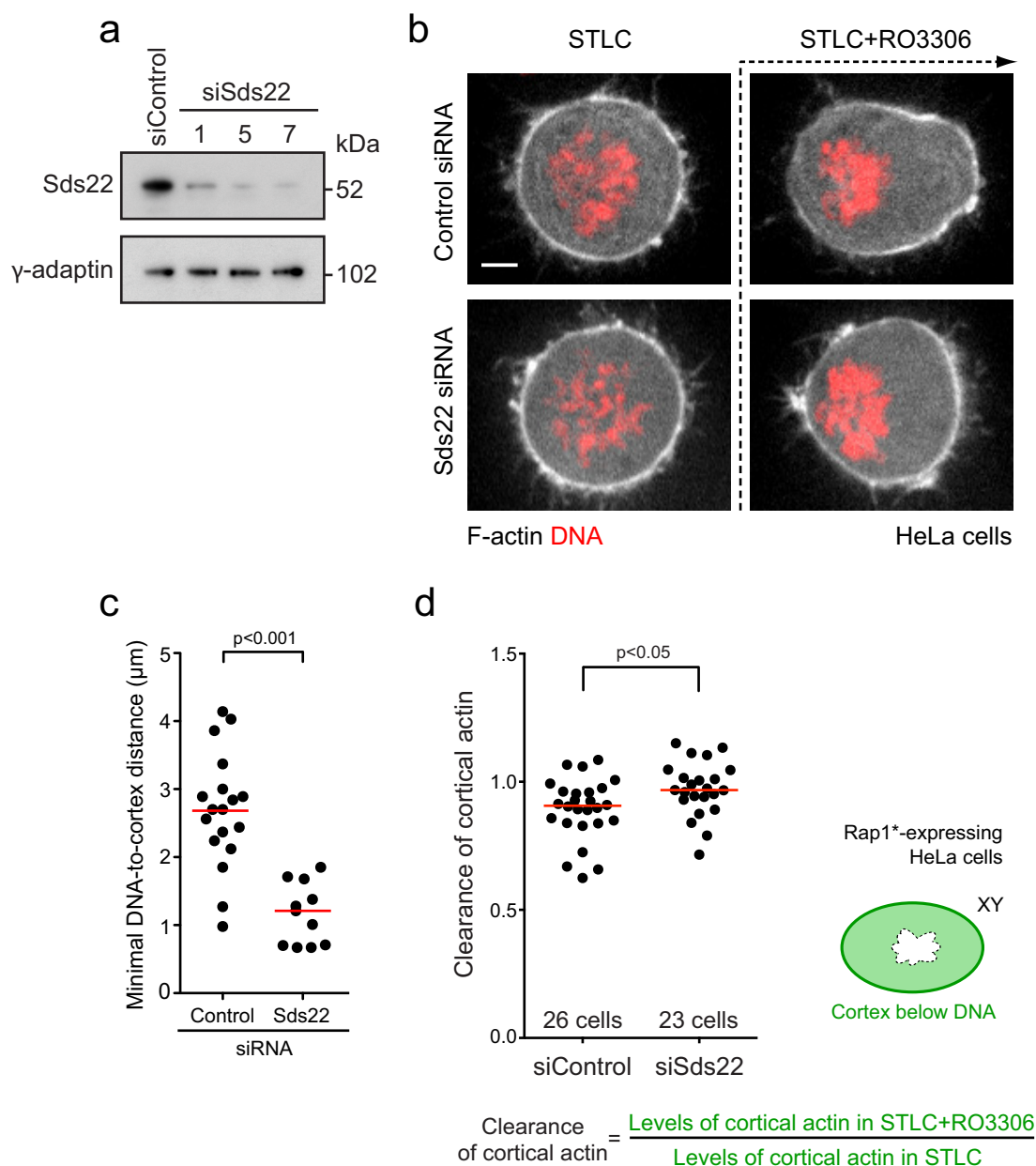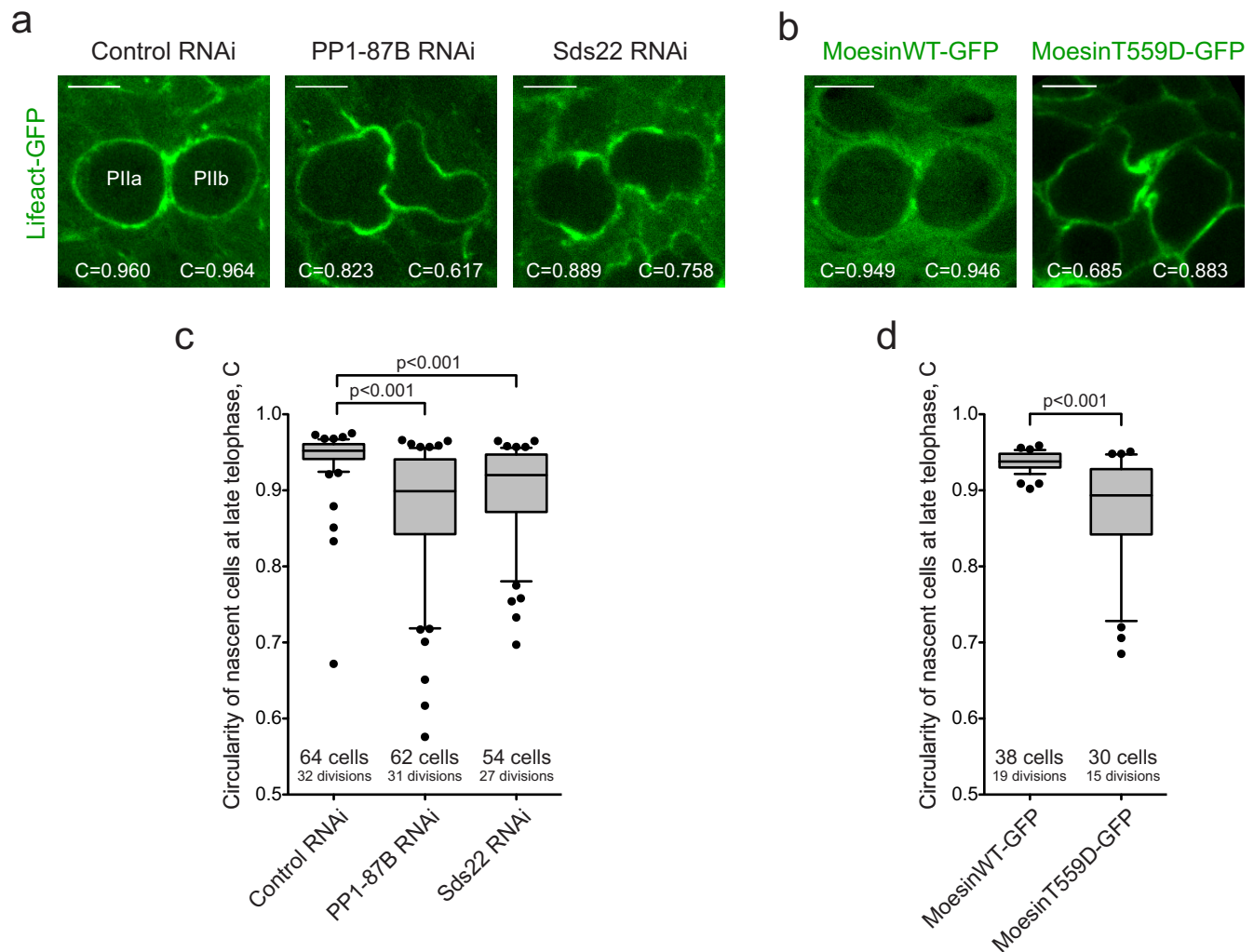
Extended Data Figure 6 | Depletion of Sds22 in human cells leads to impaired clearance of cortical actin. **a**, Western blot showing depletion of Sds22 in HeLa cells through RNA interference. **b**, Control (representative of 18 cells) and Sds22 RNAi (representative of 11 cells) STLC-treated HeLa cells expressing Lifeact–GFP and H2B–mCherry before and after RO3306 treatment (which forces cells to exit mitosis). **c**, Scatter plot (median, red line) quantifying the minimal DNA-to-cortex distance after treatment with RO3306 in each case shown in **b**. **d**, Scatter plot (median, red line) showing cortical F-actin clearance below the DNA (as seen in Extended Data Fig. 4b); siRNA oligonucleotide no. 5 (seen in **a**) was used in experiments shown in **b**–**d** and Fig. 3c, d. In **b**, scale bar, 5 μm. Significance was assessed using a two-tailed unpaired *t*-test.

a

MoesinT559D-GFP

Meta | t1   Ana | t2

15sec

t1   t2   CellY

% of cells showing polar blebbing:
Moesin-GFP - 62% (13 cells, 4 expts)
MoesinT559D-GFP - 38% (13 cells, 4 expts)

b

Moesin-GFP (CellX)
MoesinT559D-GFP (CellY)

mD=1.49μm

mD=0.37μm

Distance DNA-cortex in anaphase (μm)

Time (sec)

c

$p < 0.05$

4 experiments
13 cells   13 cells

Mean distance DNA-cortex in anaphase, mD (μm)

Moe-GFP   MoeT559D-GFP

d

Metaphase   Anaphase

Moesin-GFP

0sec   125sec

MoesinT559D-GFP

0sec   125sec

e

0sec
85sec
125sec

0sec
85sec
125sec

f

0%

$p < 0.05$

n=13   n=13

Elongation pre-furrowing (%)

Moe-GFP   MoeT559D-GFP

g

IP: PP1-87B-GFP

+CalA

min   0   10   30   0   10   30   kDa

p-FLAG-Moesin   76

FLAG-Moesin   76

PP1-87B-GFP   76

IP: PP1-87B-GFP

siLac   siSds22

min 0   10   30   0   10   30   kDa

76

76

76

h

Sds22
PP1

PM
Moesin

P

F-actin

Extended Data Figure 7 | Moesin is a target of PP1-87B–Sds22 and controls cortical relaxation at anaphase. **a–c**, The effect of constitutively active moesin on anaphase polar relaxation. **a**, An SOP cell (1 of 13 cells) expressing constitutively active moesin (moesin(T559D)) imaged in metaphase and anaphase (top panel) and a kymograph of a cross-section over time (yellow box). Only the anterior pole is indicated. **b**, Plot of the DNA-to-cortex distance over time for a representative moesin–GFP-expressing cell (cell X) and a moesin(T559D)–GFP-expressing cell (cell Y in **a**). mD, mean distance during anaphase. **c**, Box-and-whisker plot (median and 10–90 percentiles) of mD in moesin–GFP and moesin(T559D)–GFP-expressing cells. This shows that the DNA comes into close apposition to the cortex in cells expressing constitutively active moesin as the result of failure to trigger efficient polar relaxation, as it does in cells depleted for PP1-87B or Sds22. **d–f**, The effect of moesin(T559D) expression on pre-furrow elongation in the same experiment as **a–c**. **d**, Images show representative SOP cells expressing moesin–GFP or moesin(T559D)–GFP transgenes at metaphase and anaphase (out of 13 cells in each case). **e**, Outlines of the boundary of cells shown in **d** at different times during anaphase. **f**, Bar graph of cell elongation in these two backgrounds showing mean and s.d. As observed in PP1-87B- or Sds22-depleted cells, moesin(T559D)–GFP-expressing cells show aberrant cell elongation at anaphase. **g**, Immunoprecipitation (IP) assays showing moesin dephosphorylation by PP1-87B–Sds22. Calyculin A (CalA) is an inhibitor of PP1 activity (left panel). Upon addition of CalA, PP1-87B activity is suppressed, leading to higher levels of phosphorylated moesin than in the absence of the compound (see p-Flag–moesin immunoblotting). PP1-87B acts with Sds22 to dephosphorylate active moesin (right panel). Upon Sds22 depletion, PP1-87B is less efficient in inactivating moesin. Red arrows indicate PP1-87B–GFP band. Results in **g** were replicated three times. **h**, Scheme of PP1–Sds22-dependent inactivation of moesin. In **a** and **d**, scale bar, 5 μm. Significance was assessed using a two-tailed unpaired t-test. PM, plasma membrane.

Extended Data Figure 8 | Depletion of PP1-87B and Sds22 or expression of moesin(T559D)–GFP lead to severe shape defects in telophase cells.
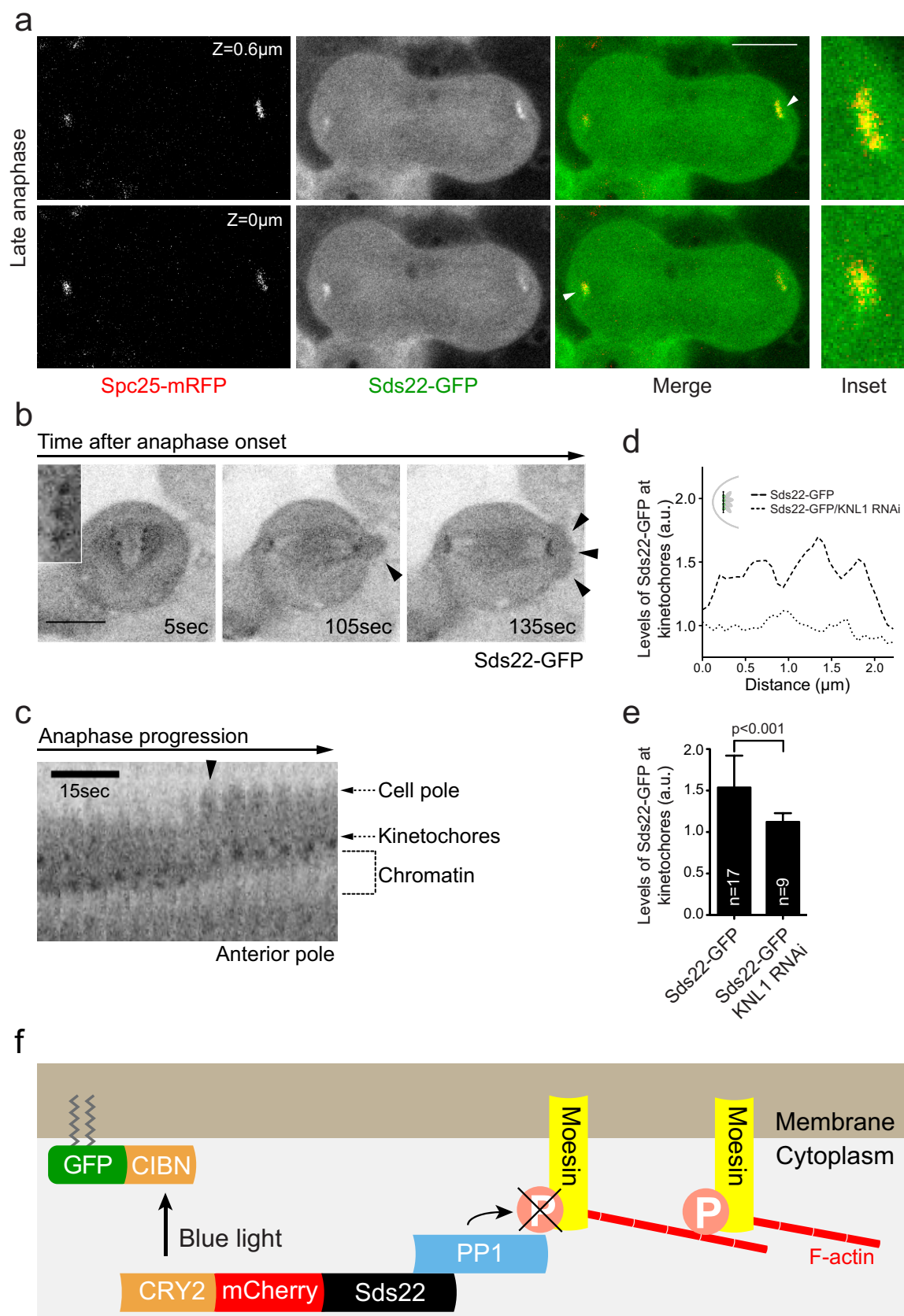**a–d**, Data show the impact of PP1-87B or Sds22 silencing and moesin(T559D) overexpression on telophase cell shape. **a**, Stills show representative telophase cells in control (1 from 32), PP1-87B RNAi (1 from 31) and Sds22 RNAi (1 from 27) backgrounds. Circularity of cells, *C*, is indicated. F-actin is labelled by Lifeact–GFP. **b**, Images show representative stills of telophase moesin–GFP (1 from 19) or moesin(T559D)–GFP (1 from 15) cells. Circularity of cells, *C*, is indicated. PIIa and PIIb are the cells that result from an asymmetric SOP division (in **a** and **b**; in each image, PIIa is the daughter cell on the left and PIIb is the daughter cell on the right.) **c**, Box plot of circularity of nascent cells at telophase in control, PP1-87B RNAi and Sds22 RNAi tissues. **d**, Box plot of circularity of nascent cells at telophase in moesin–GFP and moesin(T559D)–GFP-expressing tissues. In **a** and **b**, scale bar, 5 μm. Box-and-whisker plots show median and 10–90 percentiles. Significance was assessed using a two-tailed unpaired *t*-test.
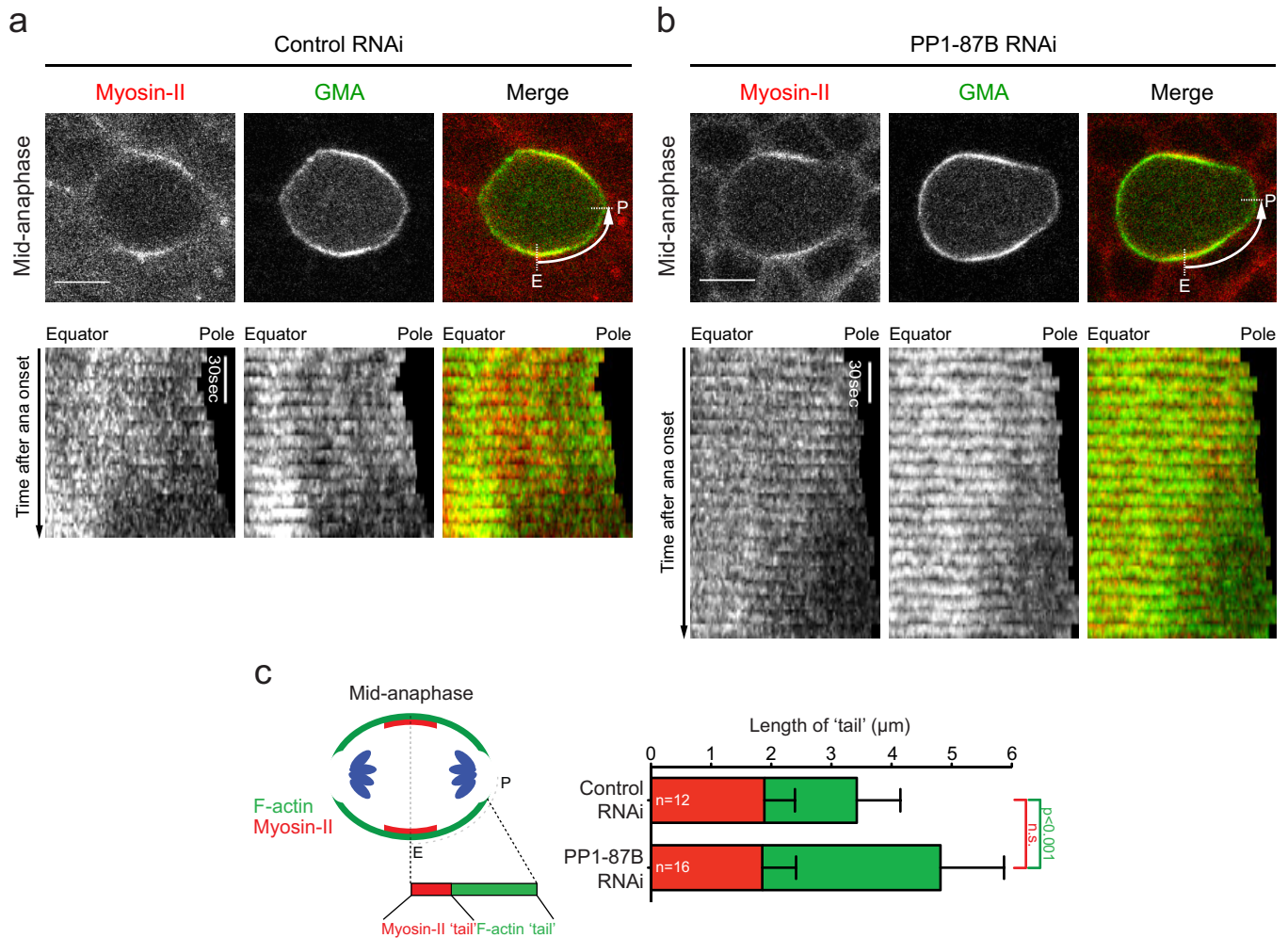
Extended Data Figure 9 | Polarization of cortical myosin II in anaphase does not depend on PP1 phosphatase. a–c, Data show the impact of PP1-87B silencing (16 cells from four animals) on myosin repolarization during anaphase onset relative to a control (12 cells from four animals). **a, b,** Stills of representative control (**a**) and PP1-87B-depleted (**b**) SOP cells in anaphase labelled for myosin II (Sqh–mCherry) and F-actin (GMA) (top), together with the corresponding kymographs showing the equator–pole (E–P) perimeter

section during anaphase progression. **c,** Schematic and graph show length of actin and myosin-II domains along the E–P perimeter in control and PP1-87B RNAi SOP cells. Mean and s.d. are shown. Significance was assessed using a two-tailed unpaired *t*-test. $P > 0.05$ was deemed not significant (n.s.). Scale bar, 5 μm. *n*, number of cells. These data show that although PP1-87B controls the polarization of cortical actin in anaphase, it does not affect the timely accumulation of myosin II at the equator.

a

Late anaphase

Z=0.6µm

Z=0µm

Spc25-mRFP   Sds22-GFP   Merge   Inset

b

Time after anaphase onset

5sec   105sec   135sec

Sds22-GFP

c

Anaphase progression

15sec

Cell pole
Kinetochores
Chromatin

Anterior pole

d

Levels of Sds22-GFP at kinetochores (a.u.)

--- Sds22-GFP
···· Sds22-GFP/KNL1 RNAi

Distance (µm)

e

Levels of Sds22-GFP at kinetochores (a.u.)

p<0.001

n=17   n=9

Sds22-GFP   Sds22-GFP KNL1 RNAi

f

GFP CIBN

Blue light

CRY2 mCherry Sds22

PP1

Moesin

Moesin

Membrane
Cytoplasm

F-actin

Extended Data Figure 10 | Local accumulation of Sds22 triggers polar blebbing in anaphase. a, Confocal cross-sections of a representative anaphase epithelial cell showing co-localization of Sds22 and the kinetochore protein Spc25 (1 of 10 cells). Insets are of regions shown by arrowheads. **b–e**, Data to assess the impact of KNL1 silencing on Sds22–GFP localization and polar relaxation. **b**, Representative epithelial cell expressing Sds22–GFP imaged during anaphase, together with a corresponding kymograph of anaphase progression in **c**. Black arrowheads point to polar blebbing. Inverted lookup table in **b**, **c**. Darker tone indicates a stronger GFP signal. In **a** and **b**, scale bar, 5 μm. **d**, Line scans across kinetochore regions denoted by the green arrowheads in representative images shown in Fig. 4a. **e**, Levels of Sds22–GFP at kinetochores in control (17 cells from four animals) and KNL1 RNAi cells (nine cells from three animals) normalized against the cytoplasmic GFP signal. Graphs show mean and s.d. Significance was assessed using a two-tailed unpaired *t*-test. **f**, Graphic representation of the blue-light-induced cryptochrome-based protein–protein interaction system underlying the data shown in Fig. 4d. This scheme shows how the CRY2-tagged Sds22 subunit interacts with membrane-tethered CIBN upon blue-light irradiation, promoting fast translocation of the phosphatase to the plasma membrane and inactivation of cortical moesin and, consequently, abrogation of F-actin linkage to the membrane.

LETTER

Structural basis for stop codon recognition in eukaryotes

Alan Brown[1]*, Sichen Shao[1]*, Jason Murray[1], Ramanujan S. Hegde[1] & V. Ramakrishnan[1]

Termination of protein synthesis occurs when a translating ribosome encounters one of three universally conserved stop codons: UAA, UAG or UGA. Release factors recognize stop codons in the ribosomal A-site to mediate release of the nascent chain and recycling of the ribosome. Bacteria decode stop codons using two separate release factors with differing specificities for the second and third bases[1]. By contrast, eukaryotes rely on an evolutionarily unrelated omnipotent release factor (eRF1) to recognize all three stop codons[2]. The molecular basis of eRF1 discrimination for stop codons over sense codons is not known. Here we present cryo-electron microscopy (cryo-EM) structures at 3.5–3.8 Å resolution of mammalian ribosomal complexes containing eRF1 interacting with each of the three stop codons in the A-site. Binding of eRF1 flips nucleotide A1825 of 18S ribosomal RNA so that it stacks on the second and third stop codon bases. This configuration pulls the fourth position base into the A-site, where it is stabilized by stacking against G626 of 18S rRNA. Thus, eRF1 exploits two rRNA nucleotides also used during transfer RNA selection to drive messenger RNA compaction. In this compacted mRNA conformation, stop codons are favoured by a hydrogen-bonding network formed between rRNA and essential eRF1 residues that constrains the identity of the bases. These results provide a molecular framework for eukaryotic stop codon recognition and have implications for future studies on the mechanisms of canonical and premature translation termination[3,4].

Termination of translation in eukaryotes is initiated when a ternary complex of eRF1–eRF3–GTP binds to a stop codon in the ribosomal A-site[5,6]. GTP hydrolysis by eRF3 induces a conformational change

that leads to its dissociation, permitting eRF1 to accommodate fully in the A-site. This change is thought to bring a universally conserved GGQ motif close to the ester bond between the nascent polypeptide and the tRNA, stimulating its hydrolysis. Concomitant with these events, the ATPase ABCE1 is recruited to the ribosome after eRF3 dissociation and, together with eRF1, catalyses splitting of the ribosomal subunits to recycle post-termination complexes[3,4,7,8].

We reasoned that a catalytically inactive eRF1 mutant may trap a pre-hydrolysis termination complex with two key features. First, eRF1 would be in complex with the stop codon it had recognized. Second, the unreleased nascent polypeptide would provide a unique affinity handle to enrich this species for structural analysis. Therefore, we substituted the glycines of the GGQ motif with alanines (eRF1(AAQ))[9] and added this mutant to in vitro translation reactions in rabbit reticulocyte lysate. Peptide release at all three stop codons was inhibited by eRF1(AAQ) as judged by persistence of a peptidyl-tRNA (Extended Data Fig. 1a, b). Affinity purification of these ribosome-nascent chain complexes (RNCs) via the nascent chain recovered both eRF1(AAQ) and ABCE1 (Extended Data Fig. 1c, d), suggesting that eRF1(AAQ) was trapped on the RNCs in its accommodated state. Association of ABCE1 was enhanced with eRF1(AAQ)-stalled RNCs relative to RNCs stalled with a truncated mRNA (Extended Data Fig. 1c), consistent with a report that the function of ABCE1 in post-termination recycling requires peptidyl-tRNA hydrolysis[7].

Purified RNCs stalled with eRF1(AAQ) at each stop codon (Extended Data Fig. 1d) were directly used for cryo-EM. Multiple rounds of three-dimensional classification in silico revealed that ~10% of the particles contained eRF1(AAQ)–ABCE1 (Extended Data Fig. 2).
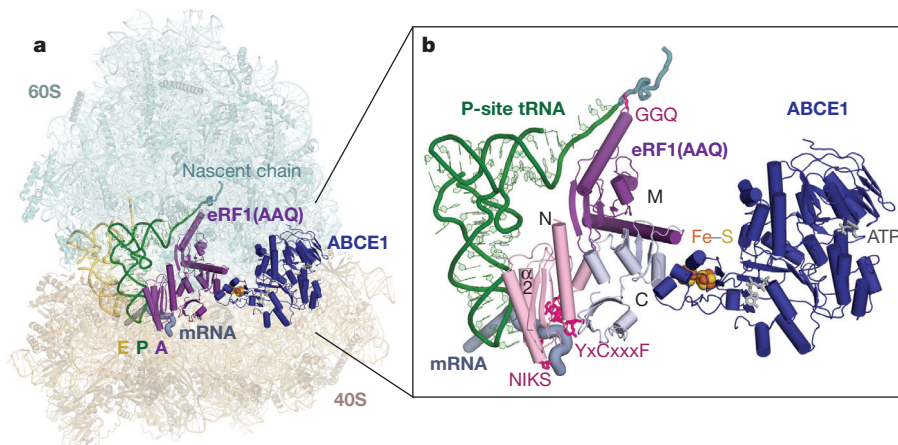

Figure 1 | Overall structure of a eukaryotic translation termination complex. a, Overview of the eRF1(AAQ)-stalled mammalian ribosome-nascent chain complex containing the UAG stop codon. Displayed are the 40S and 60S ribosomal subunits, E- and P-site tRNAs, eRF1(AAQ) occupying the A-site, and ABCE1 occupying the GTPase centre. **b,** Close-up view of

eRF1(AAQ) coloured by domain (N, M, C) with the GGQ, NIKS and YxCxxxF motifs highlighted (pink). Also shown are the P-site tRNA, nascent polypeptide, the mRNA containing the UAG stop codon, and ABCE1 with its iron–sulfur clusters and nucleotide-binding sites indicated.

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.
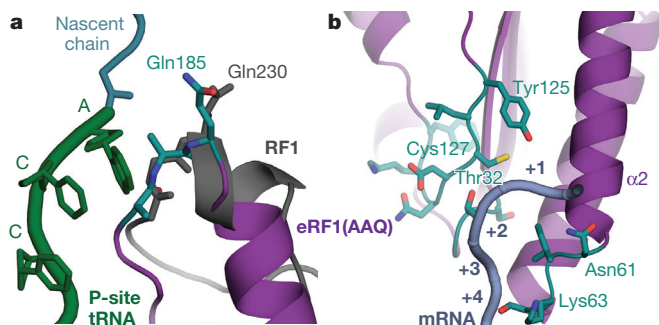*These authors contributed equally to this work

Figure 2 | Conformation of essential eRF1 motifs. a, Conformation of the GGQ-loop (teal) of eRF1(AAQ) (purple) within the peptidyl transferase centre. The AAQ tripeptide, positioned next to the CCA end of the P-site tRNA (green), closely resembles the conformation adopted by GGQ of a bacterial release factor (RF1, grey) bound to the ribosome. **b,** Positions of NIKS, YxCxxxF, and GTS motifs in the N domain of eRF1(AAQ) (purple) relative to the mRNA (slate). The positions of the stop codon (+1 to +3) and the following base (+4) are indicated.

Data sets of between 20,000 and 50,000 particles for the three stop codons yielded maps with overall resolutions of 3.65 Å (UAA), 3.45 Å (UAG) and 3.83 Å (UGA), against which the models were refined (Fig. 1, Extended Data Fig. 3 and Extended Data Table 1).

In each reconstruction, eRF1 is in its extended conformation[10], and ABCE1 occupies the GTPase centre (Fig. 1a). The three domains of eRF1 (N, M and C) have moved relative to one another compared to the crystal structure[11] (Extended Data Fig. 4a) and are each well resolved (Extended Data Fig. 3b). Direct interactions of the N domain with the codon deep in the decoding centre argues against an earlier suggestion that eRF1 disengages from the stop codon in the presence of ABCE1 (ref. 10).

The N and M domains of eRF1 independently contact the P-site tRNA and together structurally resemble a tRNA in the A-site (Fig. 1b). Helix α2 of the N domain runs parallel to, and interacts with, the anticodon stem-loop of the P-site tRNA (Extended Data Fig. 4b). The M domain is functionally analogous to the tRNA acceptor stem[11], and positions the GGQ motif[9] in the peptidyl transferase centre (Fig. 1b). To occupy a similar position as the 3′ end of an A-site tRNA, the GGQ-loop is shifted by 10 Å compared to the crystal struc-

ture[11] (Extended Data Fig. 4c). This positions the alanine residues of the mutated GGQ motif directly opposite the terminal 3′ adenosine of the P-site tRNA and the glutamine (Gln185) in proximity to the ester bond between the nascent polypeptide and the tRNA. This conformation closely resembles that of the GGQ-loop in bacterial release factors[12–14] (Fig. 2a).

At the decoding centre, the interactions between the N domain of eRF1 and the stop codon are well resolved (Fig. 2b and Extended Data Fig. 5). The most striking feature is a compact configuration of mRNA that accommodates four nucleotides in the A-site instead of three (Fig. 3a, b). mRNA compaction depends crucially on the 'flipping out' of A1825 (A1493 in bacteria) in helix 44 (h44) of 18S rRNA, forming a stacking interaction with the +2 nucleotide of the stop codon, on which the +3 nucleotide stacks. This configuration allows the +4 nucleotide to stack with base G626 (G530 in bacteria) of 18S rRNA (Fig. 3b). Stacking with G626 would be more stable for purines, explaining their statistical bias at the +4 position in eukaryotes[15]. Compaction of mRNA in the A-site probably results in pulling downstream mRNA further into the mRNA channel. This is consistent with protection of two additional nucleotides of 3′ mRNA upon eRF1 binding to the ribosome[16,17].

The structures of the bacterial release factor complexes are qualitatively distinct (Fig. 3c). Instead of A1493, the adjacent A1492 is flipped out[12–14]. However, this flipped base does not stack with any of the stop codon bases and does not lead to mRNA compaction. G530 therefore stacks with the +3 base of the stop codon[12–14] instead of the +4 base. While the +4 base also affects stop codon recognition in bacteria, the greatest preference seems to be for uridine rather than for purines[18]. Thus, eukaryotes seem to exploit the +4 base to stabilize mRNA compaction. The unique stacking of the +2 and +3 bases in this compacted state is an important element for stop codon recognition by eRF1 (see below).

The conserved NIKS, YxCxxxF and GTS motifs in eRF1 have been crosslinked to stop codon bases[19,20]. In the structures here, the essential NIKS sequence (residues 61–64)[21], located at the end of helix α2, imposes a requirement for uridine in the first (+1) position via interactions with its Watson–Crick edge. A local distortion in helix α2 and the subsequent loop helps to optimise hydrogen-bond interactions with the nucleotide (Fig. 4a). The side chains of Asn61 and Lys63 are within hydrogen-bonding distance of the uracil carbonyl groups. A further hydrogen bond may form with the main chain carbonyl of
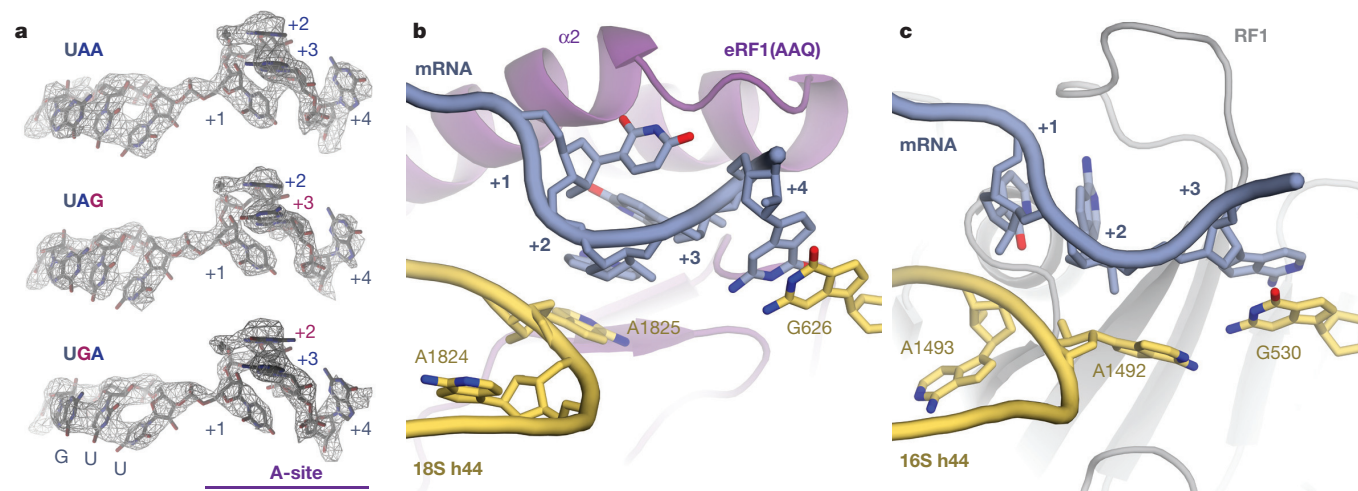

Figure 3 | Stop codon configuration in the eukaryotic decoding centre. a, Electron microscopy map densities (contoured at 0.12–0.17 e Å^{-3}) of the mRNA in each termination complex reveal the same compacted conformation. The P-site Val^{GUU} codon and the stop codon (+1 to +3) and following base (+4) in A-site are indicated. **b,** eRF1(AAQ) (purple) recognizes

four mRNA bases (+1 to +4, slate) in the A-site. Bases +2 and +3 stack on A1825, which is flipped out of helix 44 (h44), and base +4 on G626 of 18S rRNA. **c,** In bacteria, RF1 (grey) recognizes a more extended stop codon configuration where the +3 base stacks on G530 (the equivalent of G626) of 16S rRNA.
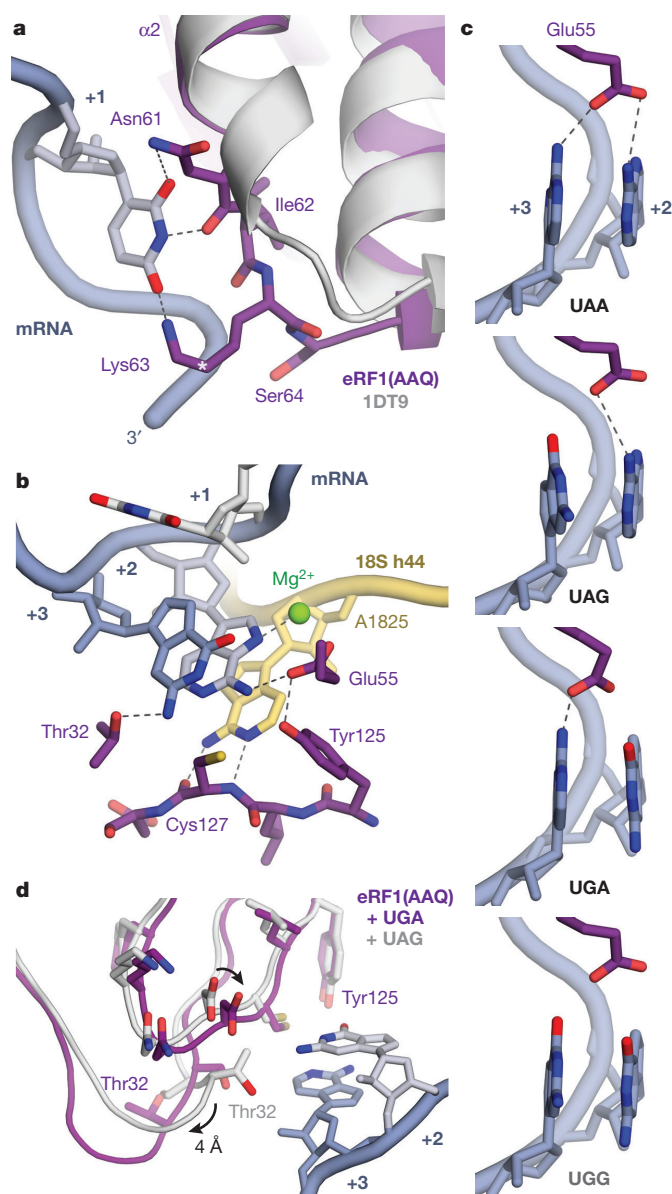
Figure 4 | Molecular basis of stop codon recognition by eRF1.
a, Comparison of ribosome-bound eRF1(AAQ) (purple) relative to the eRF1 crystal structure (PDB accession code 1DT9, light grey) shows how the NIKS motif hydrogen bonds with the +1 uridine (slate) of stop codons. Hydroxylation of Lys63 occurs on Cδ (*). **b**, Interaction network of the UAG stop codon (slate) with eRF1(AAQ) (purple) and A1825 of 18S rRNA (yellow). Interactions proposed to stabilize the flipped out position of A1825 and its stacking with the +2 and +3 bases are indicated by dotted lines. **c**, Model for stop codon discrimination by eRF1. Glu55 of eRF1(AAQ) (purple) would hydrogen bond with the stacked +2 and +3 bases (slate) of stop codons but not the sense codon UGG. **d**, Comparison of UGA-bound eRF1(AAQ) (purple) relative to UAG-bound eRF1(AAQ) (white).

Asn61. Hydroxylation of Lys63 (at the Cδ position) reduces stop codon read-through and promotes peptide release *in vivo*[22]. The hydroxyl group may allow an additional interaction with the phosphate backbone of the mRNA and help position the ε-amino group for optimal hydrogen bonding.

Purines would be disfavoured at the +1 position by steric hindrance with eRF1, while cytidine would be incompatible with the hydrogen-bonding requirements of the NIKS motif (Extended Data Fig. 6). Thus, the universality of uridine in the first position of stop codons is defined by extensive hydrogen bonding with eRF1. In bacteria, RF1 and RF2 also utilize hydrogen-bonding networks to specify

uridine in the +1 position, but with different interactions[12], highlighting how independent solutions have evolved for the same specificity problem.

Interactions of the YxCxxxF motif and Glu55 of eRF1 with the +2 and +3 purines provide a basis for stop codon discrimination from sense codons (Fig. 4b, c). The main chain of Cys127 from the YxCxxxF motif can form two hydrogen bonds with the Watson–Crick edge of A1825 to stabilize its flipped orientation, facilitating stacking of the +2 and +3 bases (Fig. 4b). The invariant residues Glu55 from helix α2 and Tyr125 may act as a pair[23] to position the glutamate side chain so that it can hydrogen bond with the N6 atoms of adenosines at the +2 and/or +3 positions. These interactions are only possible with purines, excluding pyrimidines from the +2 and +3 positions in stop codons. In the case of UGG, which codes for tryptophan, the two O6 atoms of consecutive stacked guanosines would not satisfy the hydrogen-bonding requirements, and their increased repulsion with each other and with the negatively-charged Glu55 would also disfavour the formation of the conformation that allows these interactions (Fig. 4c).

The conserved GTS motif, located towards the sugar edge of the +3 base, adopts two different conformations that are interdependent with the position of the YxCxxxF motif[24]. With adenosine in the +2 position, Thr32 faces the +3 base and can hydrogen bond with the N2 atom of the guanosine in UAG (Fig. 4b). A guanosine at the +2 position is accommodated by a movement of the YxCxxxF motif towards the stacked pair by ~1 Å, which propagates a 4 Å movement of the GTS motif so that Thr32 now faces away from the stop codon (Fig. 4d). Consistent with these observations, perturbation of any residue contributing to this network has substantial effects on stop codon recognition and specificity[17,23,25,26].

In conclusion, these structures show how stop codons are specifically selected by eRF1. At the +1 position, only uridine can form the network of interactions with the NIKS motif. The flipping of A1825 results in its stacking onto the +2 and +3 bases of a distorted mRNA so that they are decoded as a single unit (Fig. 4b). This solves the puzzle of how guanosine can occur at either the +2 or +3 position but not at both: two successive guanosines would lead to repulsion between their O6 atoms and with Glu55 (Fig. 4c). By requiring two purines while specifically excluding consecutive guanosines, the selectivity of eRF1 at the +2 and +3 positions is logically equivalent to a NAND gate.

This mechanism for stop codon recognition by eRF1 is distinct from that used by the evolutionarily and structurally unrelated bacterial release factors, even though they share some common strategies including the extensive use of hydrogen bonding for decoding specificity and an invariant GGQ motif to catalyse peptide hydrolysis. The work here paves the way for studies on the mechanism of termination and ribosome recycling, including the roles of eRF3 and ABCE1, as well as the interaction of termination factors with quality control pathways such as nonsense-mediated mRNA decay[27]. Finally, insight into stop codon recognition can provide a framework for the development of inhibitors of termination that may be useful to treat the ~11% of hereditary diseases caused by premature termination[28,29].

1. Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. Release factors differing in specificity for terminator codons. *Proc. Natl Acad. Sci. USA* **61**, 768–774 (1968).
2. Frolova, L. *et al.* A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* **372**, 701–703 (1994).
3. Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4**, a013706 (2012).
4. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. Termination and post-termination events in eukaryotic translation. *Adv. Protein Chem. Struct. Biol.* **86**, 45–93 (2012).
5. Muhs, M. *et al.* Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket paralysis virus IRES. *Mol. Cell* **57**, 422–432 (2015).

6. Taylor, D. *et al.* Cryo-EM structure of the mammalian eukaryotic release factor eRF1–eRF3-associated termination complex. *Proc. Natl Acad. Sci. USA* **109,** 18413–18418 (2012).
7. Pisarev, A. V. *et al.* The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol. Cell* **37,** 196–210 (2010).
8. Shoemaker, C. J. & Green, R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl Acad. Sci. USA* **108,** E1392–E1398 (2011).
9. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5,** 1014–1020 (1999).
10. Preis, A. *et al.* Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* **8,** 59–65 (2014).
11. Song, H. *et al.* The crystal structure of human eukaryotic release factor eRF1— mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100,** 311–321 (2000).
12. Laurberg, M. *et al.* Structural basis for translation termination on the 70S ribosome. *Nature* **454,** 852–857 (2008).
13. Weixlbaumer, A. *et al.* Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322,** 953–956 (2008).
14. Korostelev, A. *et al.* Crystal structure of a translation termination complex formed with release factor RF2. *Proc. Natl Acad. Sci. USA* **105,** 19684–19689 (2008).
15. Brown, C. M., Stockwell, P. A., Trotman, C. N. & Tate, W. P. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res.* **18,** 6339–6345 (1990).
16. Shirokikh, N. E. *et al.* Quantitative analysis of ribosome–mRNA complexes at different translation stages. *Nucleic Acids Res.* **38,** e15 (2010).
17. Kryuchkova, P. *et al.* Two-step model of stop codon recognition by eukaryotic release factor eRF1. *Nucleic Acids Res.* **41,** 4573–4586 (2013).
18. Poole, E. S., Brown, C. M. & Tate, W. P. The identity of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli. EMBO J.* **14,** 151–158 (1995).
19. Chavatte, L., Seit-Nebi, A., Dubovaya, V. & Favre, A. The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *EMBO J.* **21,** 5302–5311 (2002).
20. Bulygin, K. N. *et al.* Three distinct peptides from the N domain of translation termination factor eRF1 surround stop codon in the ribosome. *RNA* **16,** 1902–1914 (2010).
21. Frolova, L., Seit-Nebi, A. & Kisselev, L. Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA* **8,** 129–136 (2002).
22. Feng, T. *et al.* Optimal translational termination requires C4 lysyl hydroxylation of eRF1. *Mol. Cell* **53,** 645–654 (2014).
23. Kolosov, P. *et al.* Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Res.* **33,** 6418–6425 (2005).
24. Wong, L. E., Li, Y., Pillay, S., Frolova, L. & Pervushin, K. Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. *Nucleic Acids Res.* **40,** 5751–5765 (2012).
25. Cheng, Z. *et al.* Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev.* **23,** 1106–1118 (2009).
26. Seit-Nebi, A., Frolova, L. & Kisselev, L. Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep.* **3,** 881–886 (2002).
27. Czaplinski, K. *et al.* The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev.* **12,** 1665–1677 (1998).
28. Keeling, K. M., Xue, X., Gunn, G. & Bedwell, D. M. Therapeutics based on stop codon readthrough. *Annu. Rev. Genomics Hum. Genet.* **15,** 371–394 (2014).
29. Mort, M., Ivanov, D., Cooper, D. N. & Chuzhanova, N. A. A meta-analysis of nonsense mutations causing human genetic disease. *Hum. Mutat.* **29,** 1037–1047 (2008).

METHODS

No statistical methods were used to predetermine sample size and the experiments were not randomized.

Plasmids and antibodies. An SP64-based plasmid encoding 3 × Flag-tagged Sec61β containing the autonomously-folding villin headpiece (VHP) domain[30] was modified by individually inserting each stop codon (TAA, TAG, TGA) after the valine at position 68 of unmodified Sec61β. The remaining carboxy-terminal portion of the protein was deleted using Phusion mutagenesis (Thermo Scientific). *In vitro* transcription reactions were performed using PCR products generated with primers flanking the SP6 promoter and the 3′ UTR of the SP64 vector as previously described[31]. The complementary DNA encoding human eRF1 (Origene) was subcloned into a pRSETA expression vector after an amino-terminal 6 × His tag and tobacco etch virus (TEV) cleavage site using standard procedures. The eRF1(AAQ) mutant was generated via Phusion mutagenesis. Antibodies against Hbs1 and ABCE1 have been described[32]; the eRF1 antibody was from New England Biolabs.

Release factor purification. Wild-type and mutant eRF1 (eRF1(AAQ)) were expressed in and purified from *Escherichia coli* BL21(DE3) cells grown under antibiotic selection in LB. Transformed cells were induced at $A_{600\,nm} = 0.4$–0.6 with 0.2 mM IPTG for 2 h at 37 °C and lysed with a microfluidiser in lysis buffer (1 × PBS, pH 7.5, 250 mM NaCl, 10 mM imidazole, 1 mM DTT) containing 1 × protease inhibitor cocktail (Roche). Lysates were clarified by centrifugation and the supernatant passed over a NiNTA column. After washing with 25 column volumes of lysis buffer, elutions were carried out with 250 mM imidazole in lysis buffer. Peak fractions were pooled, dialysed overnight in the presence of TEV protease against 50 mM HEPES, pH 7.4, 150 mM potassium acetate, 5 mM magnesium diacetate, 10 mM imidazole, 10% glycerol, 1 mM DTT. The TEV protease and cleaved His tag were removed by passage over a NiNTA column.

***In vitro* translations and sample preparation.** *In vitro* translations in a rabbit reticulocyte (RRL) system were for 25 min at 32 °C as described previously[30,31]. Where indicated, 0.5 μM eRF1(AAQ) was included to trap termination complexes. 4 ml translation reactions were directly incubated with 100 μl (packed volume) of anti-Flag M2 beads (Sigma) for 1–1.5 h at 4 °C with gentle mixing. The beads were washed sequentially with 6 ml of 50 mM HEPES, pH 7.4, 100 mM potassium acetate, 5 mM magnesium diacetate, 0.1% Triton X-100, 1 mM DTT, 6 ml of 50 mM HEPES, pH 7.4, 250 mM potassium acetate, 5 mM magnesium diacetate, 0.5% Triton X-100, 1 mM DTT and 6 ml of RNC buffer (50 mM HEPES, pH 7.4, 100 mM potassium acetate, 5 mM magnesium diacetate, 1 mM DTT). Two sequential elutions were carried out with 100 μl 0.1 mg ml^{-1} 3 × Flag peptide (Sigma) in RNC buffer at room temperature for 25 min. The elutions were combined and centrifuged at 100,000 rpm at 4 °C for 40 min in a TLA120.2 rotor (Beckman Coulter) before resuspension of the ribosomal pellet in RNC buffer.

Electron microscopy. 3 μl aliquots of purified ribosome complexes at a concentration of 120 nM were applied onto Quantifoil R2/2 cryo-EM grids covered with continuous carbon (estimated to be 50 Å thick) at 4 °C and 100% ambient humidity. After 30 s incubation, the grids were blotted for 3 s and vitrified in liquid ethane using a Vitrobot MKIII (FEI).

Automated data collection (EPU software, FEI) was conducted on a Titan Krios microscope equipped with a XFEG electron source using 300 kV acceleration voltage. For each 1.1 s exposure, 17 movie frames were recorded on a Falcon II direct electron detector (FEI) at a calibrated magnification of 104,478, resulting in a pixel size of 1.34 Å (ref. 33). A dose rate of ∼30 electrons per Å^2 per second was used. Defocus values ranged from −1.1 to −5.9 μm for the UAA-eRF1(AAQ) data set, −0.7 to −4.1 μm for the UAG-eRF1(AAQ) data set, and −0.7 to −3.8 μm for the UGA-eRF1(AAQ) data set (Extended Data Table 1), as more images were collected closer to focus on the latter two data sets.

Image processing. The 17 frames were aligned using whole-image motion correction[34] to reduce beam-induced blurring of the images. Parameters of the contrast transfer function for each motion-corrected micrograph were obtained using Gctf (K. Zhang (kzhang@mrc-lmb.cam.ac.uk), MRC-LMB, in development). Ribosome particles were selected using semi-automated particle picking implemented in RELION[35]. All two- and three-dimensional classifications and refinements were performed using RELION[36]. We used reference-free two-dimensional class averaging to discard non-ribosomal particles and three-dimensional classification to sort different compositions and conformations of the ribosome complexes. After two-dimensional classification, the UAA-eRF1(AAQ) data set contained 556,994 particles from two independent data acquisitions, while the UAG-eRF1(AAQ) and UGA-eRF1(AAQ) data sets contained 216,276 and 250,705 particles, respectively, both from a single data acquisition. A 40 Å low-pass filtered cryo-EM reconstruction of the rabbit 80S ribosome (Electron Microscopy Data Bank (EMDB) accession 2704)[32] was used as an initial

model for the three-dimensional refinement. In a subsequent three-dimensional classification run with 12 classes, an angular sampling of 1.8° was combined with local angular searches, and the refined model from the first refinement was used as a starting model. A further round of refinement and three-dimensional classification into 6 classes using an angular sampling of 0.9° combined with local angular searches was necessary to isolate a homogeneous class of ribosomes bound with eRF1–ABCE1 (see Extended Data Fig. 2). This resulted in 49,979 particles for the UAA-eRF1(AAQ) data set, 20,515 particles for the UAG-eRF1(AAQ) data set, and 22,058 particles for the UGA-eRF1(AAQ) data set.

Prior to a final round of refinement and classification with classes combined from different data sets, statistical particle-based movie correction was performed in RELION-1.4. For these calculations we used running averages of five movie frames, and a standard deviation of one pixel for the translational alignment. In addition, we used a resolution and dose-dependent model for the radiation damage, in which each frame is B-factor weighted as estimated from single-frame reconstructions[37].

Reported resolutions are based on the Fourier shell correlation (FSC) 0.143 criterion. High-resolution noise substitution was used to correct for the effects of a soft mask on FSC curves[38]. Before visualization, density maps were corrected for the modulation transfer function of the Falcon II detector and then sharpened by applying a negative B-factor that was estimated using automated procedures[39] (Extended Data Table 1). Local resolution was quantified using ResMap[40].

Model building. The reconstruction was initially interpreted by docking the large and small subunits of the mammalian 80S ribosome, with respective Protein Data Bank (PDB) accession codes 3J92 and 3J7P[41,42], into the map using Chimera[43]. The crystal structures of human eRF1 (PDB accession code 1DT9)[11] and *Pyrococcus abysii* ABCE1 (PDB accession code 3BK7)[44] were docked into the A-site and GTPase centre, respectively, before being subjected to a Jiggle Fit in Coot[45]. To model the bound tRNAs, bacterial tRNA (PDB accession code 4V51)[46] was used as a template and modified to the sequence of the most prevalent tRNA for the particular codon from the genomic tRNA database[47]. The atomic models were then modified in Coot v0.8 to agree with the rabbit sequences and optimized for fit to density.

Model refinement and validation. Reciprocal space refinement was carried out in REFMAC v5.8 optimized for electron microscopy maps using external restraints generated by ProSMART and LIBG[45]. The model was refined against amplitudes and phases from the experimental map that were unchanged during the course of refinement. FSC$_{average}$ was monitored during refinement to follow the fit-to-density, and the final model was validated using MolProbity[48] (Extended Data Table 1). The Ramachandran statistics for eRF1 are 93.0% favoured, 1.5% outliers and for ABCE1, 91.1% favoured, 1.9% outliers. Cross-validation against over-fitting was calculated as previously described[45,49] (Extended Data Fig. 3c).

30. Shao, S., von der Malsburg, K. & Hegde, R. S. Listerin-dependent nascent protein ubiquitination relies on ribosome subunit dissociation. *Mol. Cell* **50,** 637–648 (2013).
31. Sharma, A., Mariappan, M., Appathurai, S. & Hegde, R. S. in *Protein Secretion* **619,** 339–363 (Humana Press, 2010).
32. Shao, S. & Hegde, R. S. Reconstitution of a minimal ribosome-associated ubiquitination pathway with purified factors. *Mol. Cell* **55,** 880–890 (2014).
33. Bai, X.-C., Fernandez, I. S., McMullan, G. & Scheres, S. H. W. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2,** e00461 (2013).
34. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).
35. Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189,** 114–122 (2015).
36. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
37. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3,** e03665 (2014).
38. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135,** 24–35 (2013).
39. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).
40. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11,** 63–65 (2014).
41. Voorhees, R. M., Fernandez, I. S., Scheres, S. H. W. & Hegde, R. S. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* **157,** 1632–1643 (2014).
42. Shao, S., Brown, A., Santhanam, B. & Hegde, R. S. Structure and assembly pathway of the ribosome quality control complex. *Mol. Cell* **57,** 433–444 (2015).
43. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

44. Karcher, A., Schele, A. & Hopfner, K. P. X-ray structure of the complete ABC enzyme ABCE1 from *Pyrococcus abyssi. J. Biol. Chem.* **283,** 7962–7971 (2008).
45. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71,** 136–153 (2015).
46. Selmer, M. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313,** 1935–1942 (2006).
47. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37,** D93–D97 (2009).
48. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
49. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343,** 1485–1489 (2014).
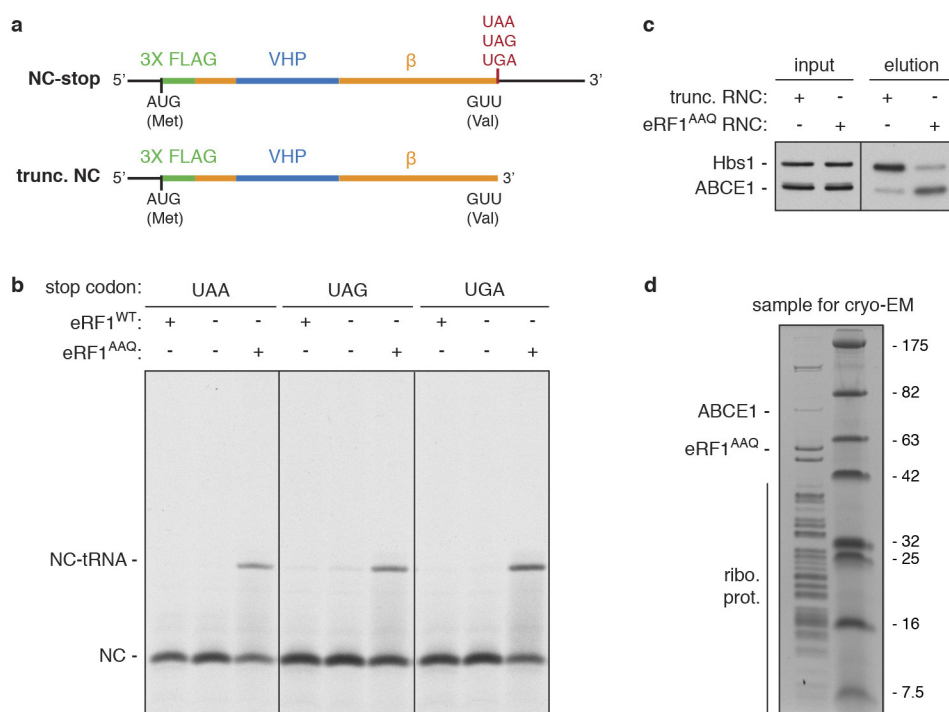
**Extended Data Figure 1 | eRF1(AAQ) stalls ribosomes at stop codons.**
a, Line diagrams of mRNA encoding nascent chain (NC) substrates used in this study. The cytosolic portion of human Sec61β (residues 1–68, orange) was modified to contain an N-terminal 3 × Flag tag (green) for affinity purification and the autonomously-folding villin headpiece (VHP, blue) domain. The three stop codons were individually inserted after Val68 of Sec61β to generate substrates for eRF1(AAQ)-mediated stalling, or the mRNA was truncated after the same residue to generate an independently-stalling substrate. **b**, *In vitro* translation reactions of NC-stop substrates containing the indicated stop codon (see panel a) in the presence of [^{35}S]methionine without or with excess eRF1 WT or eRF1(AAQ). Reactions were for 25 min at 32 °C and directly analyzed by SDS–PAGE and auto-radiography. The terminated (NC) and tRNA-associated

(NC-tRNA) nascent chain products are indicated. Addition of eRF1(AAQ) selectively prevents peptide hydrolysis when the stop codon is reached. **c**, Anti-Flag affinity purifications of ribosome-nascent chains (RNCs) stalled either by mRNA truncation or at the UAA stop codon with eRF1(AAQ) (see **a**) were immunoblotted for the splitting factors Hbs1 and ABCE1. The different amounts of Hbs1 and ABCE1 co-purified despite identical nascent chain sequences in each RNC complex suggest that eRF1(AAQ) selectively traps ABCE1 on pre-termination complexes. **d**, SDS–PAGE and Coomassie staining of affinity-purified eRF1(AAQ)-stalled ribosome-nascent chains containing the UGA stop codon used for cryo-EM analysis. Bands corresponding to ribosomal proteins, ABCE1, and eRF1(AAQ), which were verified by immunoblotting and mass spectrometry (data not shown), are indicated.
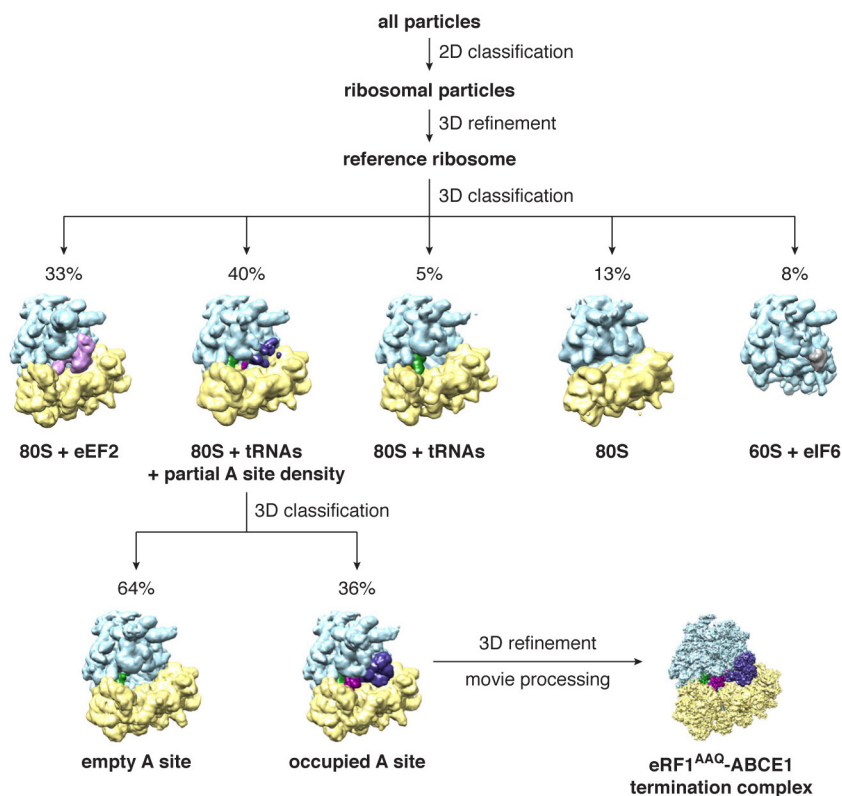
Extended Data Figure 2 | *In silico* 3D classification scheme for cryo-EM data sets. Particles extracted from automated particle picking in RELION were subjected to 2D classification. Non-ribosomal particles were discarded and the remaining particles were combined for a 3D refinement. The resulting map was used as a reference for 3D classification, which typically isolated 5 distinct classes of ribosomal complexes with the indicated distributions. Classes containing 80S ribosomes with canonical P- and E-site tRNAs and weak factor density in the A-site (~40%) were combined and subjected to another round of 3D classification for A-site occupancy. Approximately one-third of this population contained strong density for eRF1(AAQ) and ABCE1. These particles were combined for subsequent 3D refinement and movie processing. All four data sets (two for the UAA stop codon and one each for the UAG and UGA stop codons) were processed similarly. The eRF1(AAQ)–ABCE1-containing particles of the two UAA data sets after the two rounds of classification were combined for refinement to yield the final map.
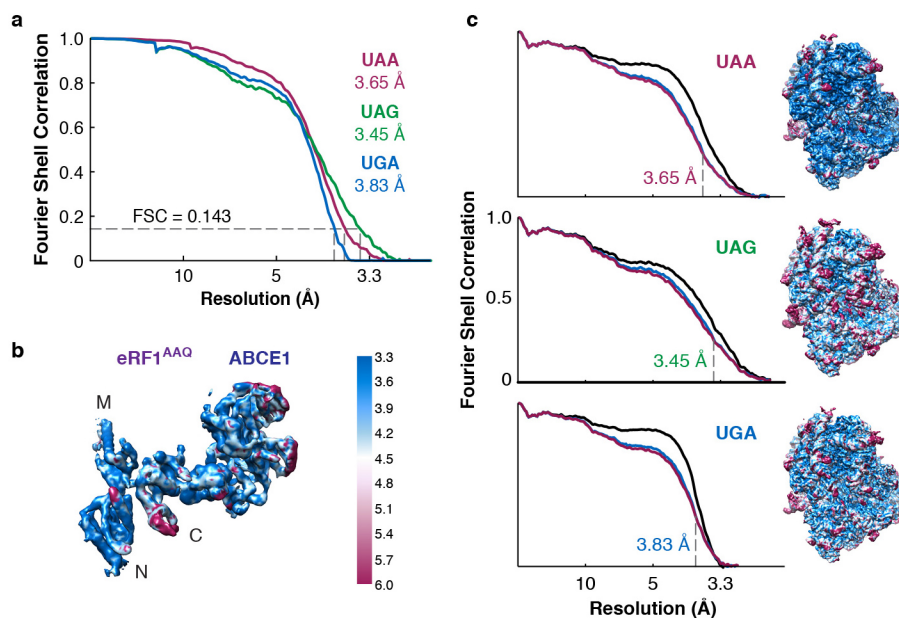
Extended Data Figure 3 | Quality of maps and models. a, Fourier shell correlation (FSC) curves for the electron microscopy maps of each termination complex containing the indicated stop codon. **b**, Isolated eRF1(AAQ)–ABCE1 density from the UAA termination complex map coloured by local resolution. **c**, Fit of models to maps. FSC curves calculated between the refined model and the final map (black), and with the self- (blue) and cross-validated (magenta) correlations for each stop codon complex. The electron microscopy map of each termination complex coloured by local resolution (as in **b**) is displayed next to the corresponding curves.
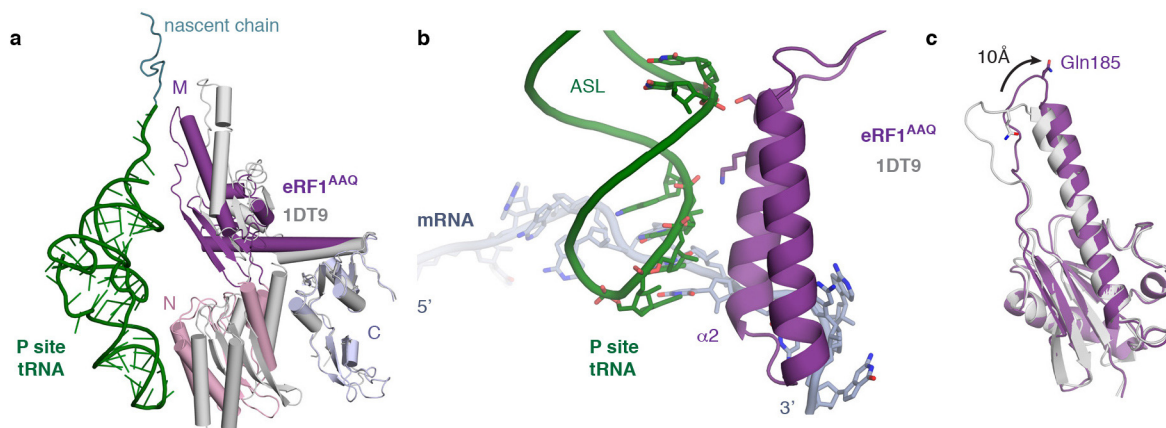
Extended Data Figure 4 | eRF1(AAQ) interactions within the termination complex. a, Comparison of ribosome-bound eRF1(AAQ) (coloured by domain) with the crystal structure of eRF1 (PDB accession code 1DT9, grey) superposed on the C domain. Both the N and M domains of eRF1 rotate upon stop codon recognition on the ribosome. The P-site tRNA (green) and nascent chain (teal) are shown for orientation. **b**, Interaction of helix α2 of the N domain of eRF1(AAQ) (purple) with the anticodon stem loop (ASL) of the P-site tRNA (green). **c**, Superposition of the eRF1(AAQ) M domain (purple) with the eRF1 crystal structure (PDB accession code 1DT9) showing a 10 Å movement of the GGQ-loop to accommodate within the peptidyl transferase centre.
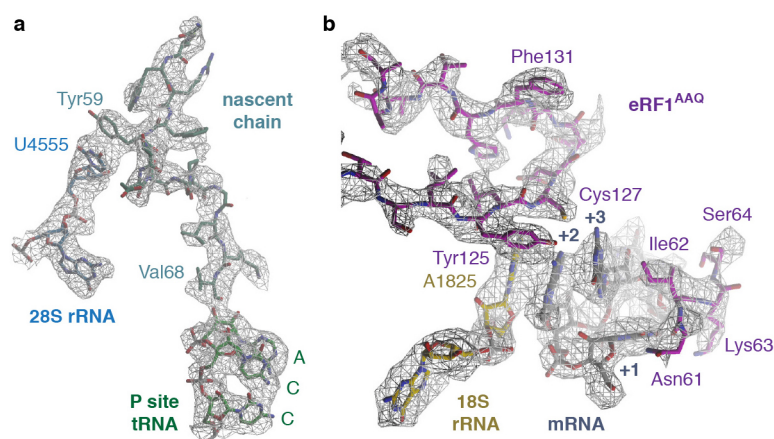
Extended Data Figure 5 | Examples of map densities. a, Density (from the UAG-containing termination complex) for the nascent chain (teal) attached to the CCA end of the P-site tRNA (green) is of sufficient resolution to model the defined sequence of the C-terminal end of the programmed nascent chain. This provides additional verification that the termination complexes are stalled at Val68 of Sec61β (human numbering) with the stop codon in the A-site (see also Extended Data Fig. 1a). A stacking interaction between an aromatic residue of the nascent chain and U4555 (blue) lining the ribosomal exit tunnel can also be observed. **b**, Densities for the interactions between the UAG stop codon (grey), a portion of h44 of 18S rRNA (yellow) and the YxCxxxF and NIKS motifs of eRF1(AAQ) (purple). The invariant isoleucine of the NIKS motif provides a hydrophobic base for the stacking of the +2 and +3 bases of the stop codon with A1825. Unlike the tyrosine and cysteine residues of the YxCxxxF motif, the phenylalanine does not contribute to stop codon recognition, but to the hydrophobic packing of the eRF1 N domain.
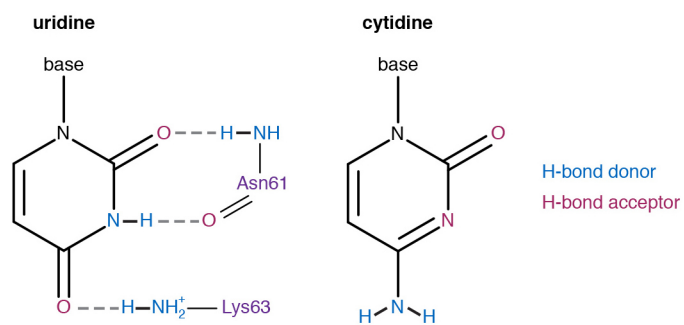
Extended Data Figure 6 | Hydrogen bonds specify for uridine at the +1 position. Chemical diagrams of uridine and cytidine with hydrogen bond donors (blue) and acceptors (magenta) indicated. Two of the three hydrogen bonds that uridine forms with Asn61 and Lys63 of the NIKS motif of eRF1(AAQ) (purple) are not possible with cytidine (see also Fig. 4a).

Extended Data Table 1 | Refinement and model statistics

	UAA	UAG	UGA
Data Collection			
Particles	49,979	20,515	22,058
Pixel size (Å)	1.34	1.34	1.34
Defocus range (μm)	1.1-5.9	0.7-4.1	0.7-3.8
Defocus mean (μm)	3.2	2.4	2.3
Voltage (kV)	300	300	300
Electron dose (e- Å^{-2})	30	30	30
Model composition			
Non-hydrogen atoms	226,532	226,533	226,533
Protein residues	12,676	12,676	12,676
RNA bases	5,820	5,820	5,820
Ligands (Zn^{2+}/Mg^{2+}/ADP)	8/197/2	8/197/2	8/197/2
Refinement			
Resolution (Å)	3.65	3.45	3.83
Map sharpening B-factor (Å^2)	-81.7	-50.6	-82.7
Average B factor (Å^2)	105.8	87.4	93.5
$FSC_{average}$	0.85	0.84	0.88
$FSC_{average}$ (eRF1)	0.70	0.64	0.74
$FSC_{average}$ (ABCE1)	0.71	0.62	0.75
R.m.s. deviations			
Bond lengths (Å)	0.006	0.006	0.008
Bond angles (°)	1.19	1.22	1.40
Validation			
Molprobity score	2.7 (93rd percentile)	2.8 (89th percentile)	3.0 (89th percentile)
Clashscore, all atoms	5.2 (100th percentile)	6.2 (97th percentile)	8.2 (97th percentile)
Good rotamers (%)	88.2	87.5	86.4
Ramachandran plot			
Favored (%)	87.0	85.8	83.5
Outliers (%)	3.3	3.5	4.2
Validation (RNA)			
Correct sugar puckers (%)	96.5	93.4	96.0
Good backbone conformations (%)	68.2	66.8	65.8

LETTER

Conformational dynamics of a class C G-protein-coupled receptor

Reza Vafabakhsh[1]*, Joshua Levitz[1]* & Ehud Y. Isacoff[1,2,3]

G-protein-coupled receptors (GPCRs) constitute the largest family of membrane receptors in eukaryotes. Crystal structures have provided insight into GPCR interactions with ligands and G proteins[1,2], but our understanding of the conformational dynamics of activation is incomplete. Metabotropic glutamate receptors (mGluRs) are dimeric class C GPCRs that modulate neuronal excitability, synaptic plasticity, and serve as drug targets for neurological disorders[3,4]. A 'clamshell' ligand-binding domain (LBD), which contains the ligand-binding site, is coupled to the transmembrane domain via a cysteine-rich domain, and LBD closure seems to be the first step in activation[5,6]. Crystal structures of isolated mGluR LBD dimers led to the suggestion that activation also involves a reorientation of the dimer interface from a 'relaxed' to an 'active' state[7,8], but the relationship between ligand binding, LBD closure and dimer interface rearrangement in activation remains unclear. Here we use single-molecule fluorescence resonance energy transfer to probe the activation mechanism of full-length mammalian group II mGluRs. We show that the LBDs interconvert between three conformations: resting, activated and a short-lived intermediate state. Orthosteric agonists induce transitions between these conformational states, with efficacy determined by occupancy of the active conformation. Unlike mGluR2, mGluR3 displays basal dynamics, which are Ca^{2+}-dependent and lead to basal protein activation. Our results support a general mechanism for the activation of mGluRs in which agonist binding induces closure of the LBDs, followed by dimer interface reorientation. Our experimental strategy should be widely applicable to study conformational dynamics in GPCRs and other membrane proteins.

Single-molecule fluorescence resonance energy transfer (smFRET) spectroscopy is a powerful tool for high-resolution probing of protein conformational change[9], and was recently applied to study membrane proteins[10–12]. To visualize ligand-induced rearrangements of full-length mGluRs, we used previously described amino-terminal SNAP- or CLIP-tagged proteins (Fig. 1a), permitting the selective and orthogonal introduction of either a FRET donor or an acceptor fluorophore into each subunit of the dimer, near the LBD[13,14]. Electrophysiological recordings in cells co-expressing the G-protein-gated inward rectifier potassium channel (GIRK) showed that these constructs were physiologically functional (Extended Data Fig. 1a). SNAP–mGluR2 and CLIP–mGluR2 were expressed in HEK293T cells and labelled with FRET donor (DY-547) and acceptor (Alexa-647) fluorophores, respectively (Methods and Extended Data Fig. 1b). Glutamate induced a concentration-dependent decrease in ensemble FRET (Extended Data Fig. 1c, d), as previously shown[15]. For the smFRET assay, we used single-molecule pull-down[16] with an anti-carboxy-terminal antibody for *in situ* immunopurification of labelled receptors from HEK293T cell lysate, followed by total internal reflection fluorescence microscopy (Fig. 1b and Extended Data Fig. 2a). The pull-down was specific, mGluR2 remained a dimer after

pull-down (Extended Data Fig. 2b, c), and there was no cross labelling between the SNAP and CLIP tags (Extended Data Fig. 2d).

In the absence of glutamate, the smFRET efficiency was ~0.45 (Fig. 1c, top), and saturating glutamate (1 mM) shifted this to ~0.2 (Fig. 1c, bottom), consistent with ensemble FRET (Extended Data Fig. 1c). Both the 0 and 1 mM glutamate states were stable within our time resolution (30 ms), with few transitions to other FRET levels. However, at intermediate glutamate concentrations, mGluR2 displayed rapid transitions between three distinct states: the 0.45 (high) FRET level seen in 0 mM glutamate, the 0.2 (low) FRET level seen in 1 mM glutamate, and a short-lived 0.35 (medium) FRET level (Fig. 1d, e and Extended Data Fig. 3a). The competitive antagonist LY341495 produced a similar FRET histogram to that seen in 0 mM glutamate: a major high FRET peak (0.45) and a minor medium FRET peak (~0.35) (Fig. 1e, bottom). About 20% of individual FRET trajectories showed visits to the low FRET state in 0 mM glutamate (Extended Data Fig. 3b), but these transitions were rare and brief and, thus, almost undetectable in the FRET histograms (Fig. 1e, top). Control experiments with an antibody against the mGluR2 N terminus instead of the C terminus showed identical histograms (Extended Data Fig. 2e, f). Moreover, the application of GTP or apyrase, to favour receptor association or dissociation from G proteins, respectively, did not alter the smFRET histograms (Extended Data Fig. 2g), indicating that G proteins are not co-immunoprecipitated with mGluR2.

Because mGluR2 did not induce G-protein signalling in the presence of 0 mM glutamate or LY341495 (Extended Data Fig. 3c), we proposed that the high and medium FRET states represent functionally inactive conformations, and that the low FRET state corresponds to the active state. Consistent with this interpretation, the low FRET state glutamate concentration-dependence had a half-maximum effective concentration (EC$_{50}$) value of 5.7 ± 0.3 μM (mean ± s.e.m.) (Fig. 1f and Extended Data Fig. 3d), corresponding to the concentration-dependence of GIRK current activation in HEK293T cells (3.2 ± 0.3 μM) (Extended Data Fig. 3e). Moreover, glutamate had no effect on the FRET histogram in the glutamate-insensitive mutant mGluR2-YADA (Tyr216Ala, Asp295Ala)[17] (Extended Data Fig. 3g). Finally, the addition of LY341495 to glutamate abolished the low FRET state (Extended Data Fig. 3h). These observations confirm the assignment of the low FRET peak to the active conformation.

We next quantified the glutamate-induced fluctuations with cross-correlation and dwell-time analyses. The cross-correlation amplitude of donor and acceptor signals increased with glutamate concentration, reaching a maximum near the EC$_{50}$ value, and decreased at high glutamate concentrations (Fig. 1g). Anti-correlation between donor and acceptor was almost abolished in saturating glutamate (1 mM), confirming the stabilization of the active state. Kinetic analysis of individual traces showed an ~84-ms active state dwell time that was nearly independent of glutamate concentration (Fig. 1h and Extended Data Fig. 3i), suggesting that the active state dwell time reflects the glutamate dissociation rate and the inherent stability of the active

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [2]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. [3]Physical Bioscience Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
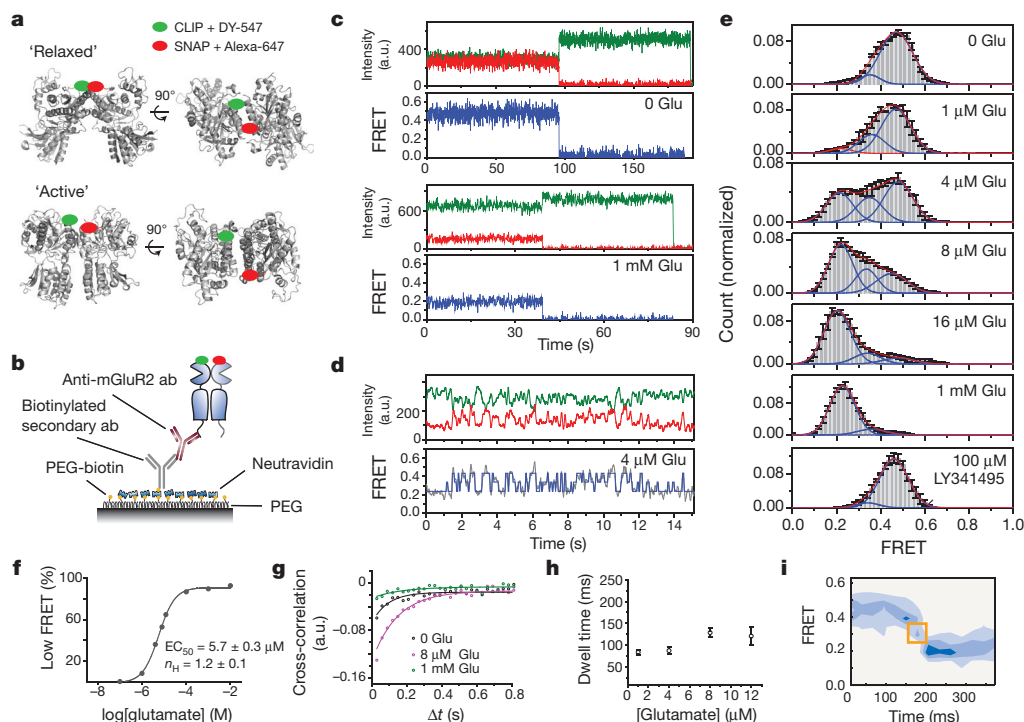*These authors contributed equally to this work.

Figure 1 | A single-molecule FRET assay reveals three conformations of the mGluR2 activation pathway. a, Crystal structures of mGluR1 in the 'relaxed' (Protein Data Bank (PDB) accession 1EWT) and 'active' states (PDB code IEWK) show an increase in the distance between N termini after activation. Red and green ovals show the approximate positions of SNAP and CLIP tags, respectively. **b**, Schematic of single-molecule FRET measurements. Ab, antibody. **c**, Donor (green) and acceptor (red) intensity time traces and FRET trace (blue) in the absence (top) or presence (bottom) of 1 mM glutamate show a decrease in FRET in the presence of saturating glutamate. a.u., arbitrary units. **d**, Representative smFRET traces at 4 μM glutamate reveal rapid dynamics between three states. A three-state fit obtained from hidden Markov analysis is overlaid over the filtered raw data. **e**, smFRET histograms in the presence of a range of glutamate concentrations or a competitive antagonist (LY341495). Blue lines show global three-component Gaussian fits that show the high (~0.45), medium (~0.35) and low (0.2) FRET states. The sum of all three components is shown in red. **f**, Titration curve for the low FRET peak. n_H, Hill coefficient. **g**, Cross-correlation plots show limited dynamics in the absence of glutamate (black) or in saturating glutamate (1 mM, green), but enhanced dynamics at intermediate concentrations (8 μM, magenta). Solid lines show single exponential fits. **h**, Concentration dependence of low FRET dwell times obtained from dwell time analysis. **i**, FRET density plots constructed from synchronized transitions from the high to low FRET states show a short dwell at the medium FRET level (yellow box). Error bars are s.e.m.

conformation. Two-dimensional histograms obtained from synchronized transitions into or out of the low FRET state showed a short dwell at the medium FRET value of ~0.35, providing additional evidence that this state is an intermediate in the activation pathway (Fig. 1i and Extended Data Fig. 3j).

Next, we investigated the relationship between ligand efficacy and receptor conformation. We studied two group II mGluR agonists, the efficacies of which differ from that of glutamate[15] (Fig. 2a). While DCG-IV increased the occupancy of the low FRET state, even at saturating concentrations (≥100 μM), the low FRET state was not fully occupied and ~30% of the distribution remained in the high and medium FRET states (Fig. 2b). Single-molecule traces in saturating DCG-IV showed recurrent transitions out of the low FRET state (Fig. 2c, d and Extended Data Fig. 4a). By contrast, the full-agonist LY379268 populated the same three FRET states but with more complete occupancy of the active state than glutamate (Fig. 2e and Extended Data Fig. 4b, c). Consistent with these results, cross-correlation analysis showed that mGluR2 retained substantial dynamics in saturating DCG-IV, but not in saturating glutamate or LY379268 (Fig. 2d). Notably, at concentrations that produced comparable occupancy of the low FRET state (Extended Data Fig. 4d), cross-correlation analysis showed that LY379268 induced longer timescale fluctuations and longer active state dwell times than did glutamate and DCG-IV (Fig. 2f–h and Extended Data Fig. 4e, f). Two-dimensional histograms showed that both DCG-IV and LY379268, like glutamate, visit the medium FRET state as an intermediate during activation (Extended Data Fig. 4g, h). Thus, unlike ionotropic receptors, in which the efficacy

of an agonist is a function of the degree of LBD closure[18], in mGluR2, agonists with different efficacies stabilize the same conformational states, with the degree of efficacy depending on the occupancy of a single active conformation. The data suggest that ligand efficacy depends more on the rate of transition into the active state than on dwell time in the active state; that is, that agonists differ in how they induce LBD closure and subsequent reorientation. Additionally, an mGluR2-specific positive allosteric modulator, LY48739, and the transmembrane domain (TMD) mutations Gln679Val and Cys770Ala (refs 19, 20), which all increase the efficacy of DCG-IV, also increased the relative occupancy of the low FRET state (Extended Data Fig. 5), further supporting the role of active state occupancy in determining agonist efficacy, and confirming that the immobilized receptors retain TMD function and coupling to the LBD.

We next wondered whether the properties of mGluR2 apply to other mGluRs. We turned to the other group II mGluR, mGluR3, which possesses ~70% sequence identity with mGluR2. SNAP- or CLIP-tagged mGluR3 constructs were physiologically functional (Extended Data Fig. 6a) and underwent a glutamate-dependent decrease in ensemble FRET similar to mGluR2, but with a lower EC_{50} value (0.5 ± 0.2 μM). Notably, single-molecule trajectories of mGluR3 in 0 mM glutamate exhibited frequent transitions between the three FRET states (Fig. 3a, top), resulting in ~30% occupancy of the low FRET active state (Extended Data Fig. 6b, c), which was eliminated by LY341495 (Fig. 3a, bottom; Extended Data Fig. 6b, c). LY341495 produced a large decrease in the glutamate-free GIRK current (Fig. 3b, c and Extended Data Fig. 6d) and ensemble FRET (Extended Data
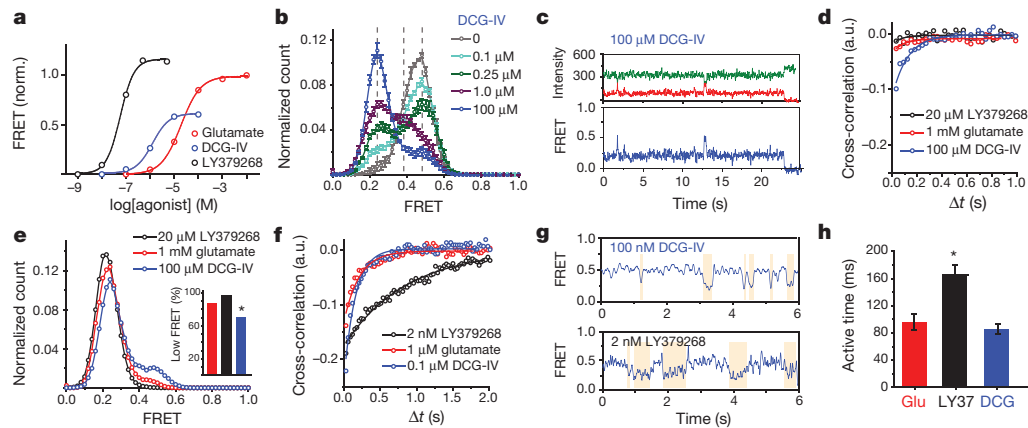
Figure 2 | Conformational basis of partial agonism of mGluR2. a, Ensemble
FRET titrations in HEK293T cells expressing mGluR2 in the presence of
glutamate (red), DCG-IV (blue) or LY379268 (black). FRET values in each
condition are normalized to the response to 1 mM glutamate. **b**, smFRET
histogram for DCG-IV shows the same three states as seen for glutamate
(dotted lines), with dose-dependent occupancy of the low FRET state.
c, Representative smFRET trace shows transitions out of the low FRET state in
saturating DCG-IV. **d**, Cross-correlation plots for saturating agonist reveal
dynamics = DCG-IV > glutamate > LY379268. Solid lines show single

exponential fits. **e**, smFRET histograms for saturating agonist; occupancy
of the low FRET state = DCG-IV < glutamate < LY379268 (inset). *$P = 0.008$,
two-tailed t-test. **f**, **g**, At concentrations that result in comparable population
of the active state, LY379268 induces slower dynamics than DCG-IV and
glutamate as shown in cross-correlation (**f**) and representative smFRET traces
(**g**). **h**, LY379268 induces significantly longer low FRET state dwell times than glutamate and DCG-IV
(two-tailed t-test, *$P = 0.0084$). Error bars are s.e.m.

Fig. 6e, f) in cells expressing mGluR3, but not in cells expressing
mGluR2, indicating that, unlike mGluR2, mGluR3 has basal activity.
Importantly, the 0 mM glutamate smFRET distribution of mGluR3
was insensitive to GTP, indicating that the different basal activity of
mGluR3 is not due to association with G protein (Extended Data
Fig. 6g).

Earlier work suggested that some mGluRs, including mGluR3 but
not mGluR2, are calcium-sensitive[21]. However, direct binding of Ca^{2+}
has been challenged, and Ca^{2+} activity was attributed to indirect
downstream signalling effects[22]. smFRET measurements of mGluR3
in 0 mM glutamate and 0 mM Ca^{2+} showed a significant reduction of
basal dynamics and occupancy of the low FRET state (Fig. 3d, e and
Extended Data Fig. 7a, b). In 0 mM glutamate, we observed a Ca^{2+}-
concentration-dependent increase in low FRET state occupancy
dynamics (Extended Data Fig. 7c, d). By contrast, removal of Ca^{2+}
did not alter the smFRET properties of mGluR2 (Extended Data

Fig. 7f, g). Saturating glutamate (0.5 mM) induced the same mGluR3
smFRET histogram in 0 and 2 mM Ca^{2+}, indicating that Ca^{2+} is not
required for full agonism of mGluR3 (Fig. 3d, e). Similar to the orthos-
teric agonists on mGluR2, in mGluR3, Ca^{2+} induced transient occu-
pancy of the intermediate FRET state (Extended Data Fig. 7e). The
mutation Ser152Asp, which is reported to abolish the Ca^{2+} sensitivity
of both mGluR1 and mGluR3 (ref. 21), eliminated the effect of
LY341495 on mGluR3 (Extended Data Fig. 8a–c), indicating elimina-
tion of basal activity. Consistent with this, smFRET analysis of
mGluR3(Ser152Asp) revealed a large reduction in basal low FRET
population and Ca^{2+} sensitivity (Fig. 3d and Extended Data Fig. 8d, e)
and a decrease in receptor dynamics (Fig. 3f and Extended Data
Fig. 8f). Dwell-time analysis of mGluR3 in the presence of 2 mM
Ca^{2+} or 100 nM glutamate (Fig. 3g, h and Extended Data Fig. 9a, b)
showed an average active state lifetime of 183 ms, significantly longer
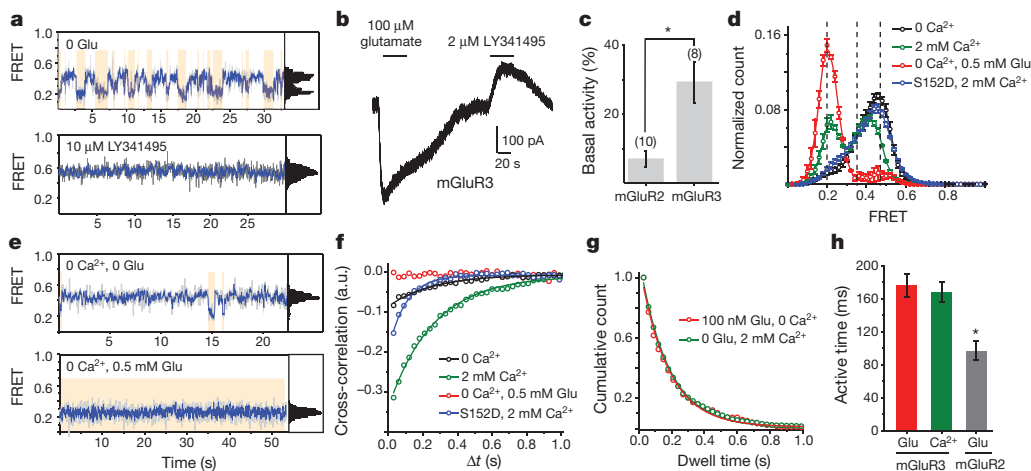than the ~80 ms lifetime in mGluR2 (Fig. 1h), indicating that the

Figure 3 | mGluR3 has high basal structural dynamics and activity.
a, Representative mGluR3 smFRET traces show basal dynamics in the absence
of glutamate (top) that are abolished by the competitive antagonist LY341495
(bottom). **b**, In HEK293T cells co-expressing GIRK channels, mGluR3 has
basal activity in the absence of glutamate, which is blocked by LY341495.
c, Basal activity ($[I_{LY341495}]/([I_{LY341495}] + [I_{Glu}])$) for mGluR2 and mGluR3.
Values in parentheses indicate number of cells tested. *$P = 0.0097$ (unpaired
two-tailed t-test). **d**, smFRET histograms for mGluR3 show occupancy of the

low FRET state that is abolished by the removal of Ca^{2+} or introduction of the
Ser152Asp mutation. **e**, Representative smFRET traces for mGluR3 in the
absence of Ca^{2+} with either 0 (top) or saturating (bottom) glutamate. **f**, Cross-
correlation plots fit to a single exponential function for mGluR3. **g**, **h**, Dwell
time of the active state for mGluR3 in the presence of glutamate (100 nM) or
Ca^{2+} (2 mM) compared to mGluR2 (4 µM glutamate) (unpaired two-tailed
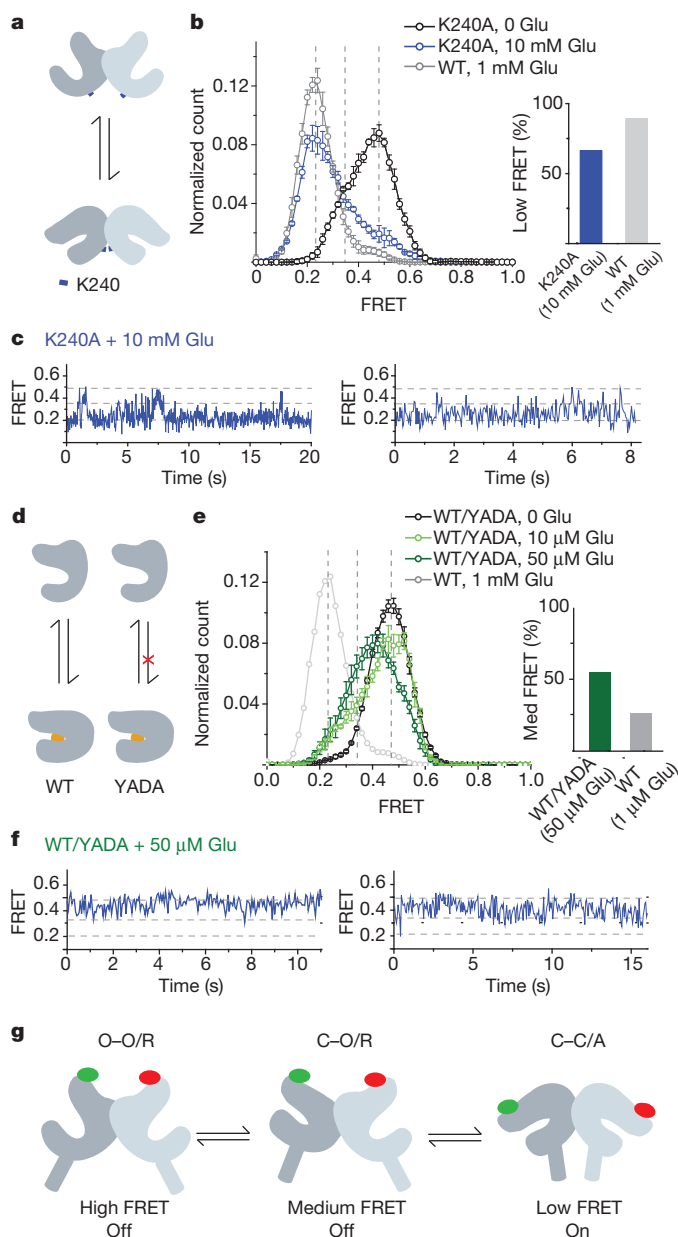t-test, *$P = 0.00055$). Error bars are s.e.m.

Figure 4 | A three-state model of mGluR activation. a, b, Mutation of residue Lys240 at the lower lobe LBD dimer interface (blue marker) (**a**) decreases occupancy of the low FRET state at saturating (10 mM) glutamate (**b**). WT, wild type. **c,** Representative smFRET traces for Lys240Ala in the presence of glutamate show transitions out of the low FRET state. **d, e,** Heterodimers of wild-type mGluR2 and the glutamate-insensitive mGluR2 (YADA) mutant (**d**) show enhanced occupancy of the medium FRET state (**e**). **f,** Representative smFRET traces for WT/YADA heterodimers show transitions between the high and medium FRET states with limited visits to the low FRET state. **g,** Three-state structural model of mGluRs based on intra-subunit (closed 'C' to open 'O') and inter-subunit (relaxed 'R' to active 'A') conformational changes. Error bars are s.e.m.

active conformation of mGluR3 is more stable than in mGluR2. Consistent with this, the time scale of donor and acceptor cross-correlation for mGluR3 in all conditions was much slower than for mGluR2 (Extended Data Fig. 9c). Overall, smFRET measurements revealed that mGluR3 transitions between the same three FRET states as mGluR2. However, unlike mGluR2, mGluR3 is Ca^{2+}-sensitive and therefore basally active under physiological conditions.

Having observed three ligand-dependent FRET states in the LBDs of both mGluR2 and mGluR3, we sought to identify the underlying

conformational rearrangements in the activation pathway. The available structures of mGluR LBDs have been characterized as either 'relaxed', with the lower lobes of the LBD far apart, or 'active', with the lower lobes of the LBD closer to one other (Fig. 4a and Extended Data Fig. 10a). We proposed that electrostatic interactions between charged residues in the lower lobe, including a conserved lysine (Lys240 in mGluR2), stabilize the active conformation. Indeed, neutralizing Lys240 (Lys240Ala) decreased the apparent affinity for glutamate in both ensemble FRET and GIRK activation assays in cells (Extended Data Fig. 10b). In the smFRET assay, even at saturating glutamate concentrations (10 mM), this mutant populated the low FRET state less than wild-type mGluR2 (Fig. 4b). Single-molecule trajectories showed frequent transitions out of the low FRET state in saturating glutamate (Fig. 4c and Extended Data Fig. 10c), consistent with destabilization of the active state. These results support the idea that the lower lobes of mGluR2 come into close proximity in the active state and help to stabilize it.

We next investigated the conformations corresponding to the high (0.45) and medium (0.35) FRET states. Considering the comparatively small distance change between high and medium FRET states (~4 Å) compared to medium and low FRET states (~8 Å), and the observation that the medium FRET state seems to be inactive, we proposed that the medium state corresponds to a relaxed conformation in which only one LBD has closed. If this is true, an mGluR2 heterodimer composed of a wild-type subunit and a YADA subunit (wild-type/YADA) is expected to bind glutamate only in the wild-type LBD, and therefore populate the middle FRET state more than the wild-type homodimer. smFRET analysis showed that at near-saturating concentrations for the wild-type subunit, wild-type/YADA had an ~55% occupancy of the medium FRET state, whereas wild-type/wild-type had a maximal medium FRET occupancy of ~25% (Fig. 4d, e and Extended Data Fig. 10d–f). Unlike the mono-phasic concentration-dependence of occupancy of the low FRET state seen in wild-type mGluR2 (Fig. 1f), wild-type/YADA showed a biphasic distribution (Extended Data Fig. 10g), resembling the previously reported dose-response of effector activation by the wild-type/YADA heterodimer of mGluR5 (ref. 19), and supporting the assignment of the low smFRET conformation to the active state. smFRET traces at 50 μM glutamate showed numerous transitions between the high and medium FRET states with rare and brief visits to the low FRET state (Fig. 4f and Extended Data Fig. 10h), consistent with the activity of this heterodimer, and possibly due to occasional binding of glutamate to the YADA subunit or spontaneous closure of the YADA subunit in the absence of glutamate.

Our kinetic, mutational and functional analyses indicate that mGluR2 and mGluR3 undergo ligand-dependent fluctuations between three conformations: a resting and inactive O–O/R conformation, an active C–C/A conformation and an intermediate inactive short-lived C–O/R conformation (in which 'C' denotes 'closed', 'O' denotes 'open', 'R' denotes 'relaxed' and 'A' denotes 'active') (Fig. 4g). The relative instability of the intermediate conformation may explain why crystal structures have not been obtained in the C–O/R state.

Electrostatic interactions at the lower lobe LBD interface stabilize the active state, suggesting that mGluR activation requires closure of both LBDs followed by rearrangement of the dimer interface (Fig. 4g) and consistent with findings that activation requires downstream reorientation at the inter-subunit interfaces between cysteine-rich domains and TMDs[23–26]. We observed occupancy times of tens of milliseconds to seconds—on the timescale of G protein signalling[27]—for both the O–O/R and C–C/A conformations. This is longer than the sub-millisecond fluctuations seen in isolated LBDs using diffusion-based FRET[28], suggesting that the TMD of the intact receptor stabilizes the LBDs in the R and A states, giving the receptors a wide dynamic range of activity. Our findings suggest that the fractional occupancy of the C–C/A conformation determines agonist efficacy, consistent with isolated LBD single-molecule spectroscopy[28] and crystal structures, in

which the degree of closure is similar for full and partial agonists[18,19,29]. Finally, we revealed kinetic differences between mGluR2 and mGluR3 and found that mGluR3 has a more stable active state and is activated by physiological concentrations of external Ca^{2+}, resulting in considerable basal G-protein signalling in cells. Our study provides a framework for investigating the activation mechanisms of other bi-lobed, clamshell LBDs, such as in the $GABA_B$ receptor and ionotropic neurotransmitter receptors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Rasmussen, S. G. et al. Crystal structure of the β₂ adrenergic receptor-Gs protein complex. Nature 477, 549–555 (2011).
2. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. Annu. Rev. Pharmacol. Toxicol. 53, 531–556 (2013).
3. Conn, P. J. & Pin, J. P. Pharmacology and functions of metabotropic glutamate receptors. Annu. Rev. Pharmacol. Toxicol. 37, 205–237 (1997).
4. Niswender, C. M. & Conn, P. J. Metabotropic glutamate receptors: physiology, pharmacology, and disease. Annu. Rev. Pharmacol. Toxicol. 50, 295–322 (2010).
5. Kniazeff, J. et al. Locking the dimeric GABAB G-protein-coupled receptor in its active state. J. Neurosci. 24, 370–377 (2004).
6. Kumar, J. & Mayer, M. L. Functional insights from glutamate receptor ion channel structures. Annu. Rev. Physiol. 75, 313–337 (2013).
7. Kunishima, N. et al. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. Nature 407, 971–977 (2000).
8. Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H. & Morikawa, K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd³⁺. Proc. Natl Acad. Sci. USA 99, 2660–2665 (2002).
9. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. Nature Methods 5, 507–516 (2008).
10. Zhao, Y. et al. Single-molecule dynamics of gating in a neurotransmitter transporter homologue. Nature 465, 188–193 (2010).
11. Bockenhauer, S., Furstenberg, A., Yao, X. J., Kobilka, B. K. & Moerner, W. E. Conformational dynamics of single G protein-coupled receptors in solution. J. Phys. Chem. B 115, 13328–13338 (2011).
12. Morrison, E. A. et al. Antiparallel EmrE exports drugs by exchanging between asymmetric structures. Nature 481, 45–50 (2011).
13. Keppler, A. et al. A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nature Biotechnol. 21, 86–89 (2003).
14. Doumazane, E. et al. A new approach to analyze cell surface protein complexes reveals specific heterodimeric metabotropic glutamate receptors. FASEB J. 25, 66–77 (2011).
15. Doumazane, E. et al. Illuminating the activation mechanisms and allosteric properties of metabotropic glutamate receptors. Proc. Natl Acad. Sci. USA 110, E1416–E1425 (2013).
16. Jain, A. et al. Probing cellular protein complexes using single-molecule pull-down. Nature 473, 484–488 (2011).
17. Kniazeff, J. et al. Closed state of both binding domains of homodimeric mGlu receptors is required for full activity. Nature Struct. Mol. Biol. 11, 706–713 (2004).
18. Jin, R., Banke, T. G., Mayer, M. L., Traynelis, S. F. & Gouaux, E. Structural basis for partial agonist action at ionotropic glutamate receptors. Nature Neurosci. 6, 803–810 (2003).
19. Yamashita, T., Kai, T., Terakita, A. & Shichida, Y. A novel constitutively active mutation in the second cytoplasmic loop of metabotropic glutamate receptor. J. Neurochem. 91, 484–492 (2004).
20. Yanagawa, M., Yamashita, T. & Shichida, Y. Activation switch in the transmembrane domain of metabotropic glutamate receptor. Mol. Pharmacol. 76, 201–207 (2009).
21. Kubo, Y., Miyashita, T. & Murata, Y. Structural basis for a Ca²⁺-sensing function of the metabotropic glutamate receptors. Science 279, 1722–1725 (1998).
22. Nash, M. S., Saunders, R., Young, K. W., Challiss, R. A. & Nahorski, S. R. Reassessment of the Ca²⁺ sensing property of a type I metabotropic glutamate receptor by simultaneous measurement of inositol 1,4,5-trisphosphate and Ca²⁺ in single cells. J. Biol. Chem. 276, 19286–19293 (2001).
23. Tateyama, M., Abe, H., Nakata, H., Saito, O. & Kubo, Y. Ligand-induced rearrangement of the dimeric metabotropic glutamate receptor 1alpha. Nature Struct. Mol. Biol. 11, 637–642 (2004).
24. Huang, S. et al. Interdomain movements in metabotropic glutamate receptor activation. Proc. Natl Acad. Sci. USA 108, 15480–15485 (2011).
25. Hlavackova, V. et al. Sequential inter- and intrasubunit rearrangements during activation of dimeric metabotropic glutamate receptor 1. Sci. Signal. 5, ra59 (2012).
26. Xue, L. et al. Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. Nature Chem. Biol. 11, 134–140 (2015).
27. Lohse, M. J., Maiellaro, I. & Calebiro, D. Kinetics and mechanism of G protein-coupled receptor activation. Curr. Opin. Cell Biol. 27, 87–93 (2014).
28. Olofsson, L. et al. Fine tuning of sub-millisecond conformational dynamics controls metabotropic glutamate receptors agonist efficacy. Nature Commun. 5, 5206 (2014).
29. Muto, T., Tsuchiya, D., Morikawa, K. & Jingami, H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. Proc. Natl Acad. Sci. USA 104, 3759–3764 (2007).

Author Contributions R.V., J.L. and E.Y.I. designed the research. R.V. set up, performed and analysed single-molecule FRET experiments. J.L. performed and analysed ensemble FRET and electrophysiology experiments and contributed to single-molecule FRET experiments. R.V., J.L. and E.IY. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.Y.I. (ehud@berkeley.edu).

METHODS

No statistical methods were used to predetermine sample size. Experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture. HEK293T cells were cultured in DMEM with 5% FBS on poly-L-lysine-coated glass coverslips. HEK293T cells were obtained from the UC Berkeley MCB tissue culture facility, authenticated by DDC Medical, and tested negative for mycoplasma contamination. Previously described HA–SNAP and Flag–CLIP-tagged rat mGluR cDNA were provided by J. P. Pin. DNA plasmids were transfected into cells using lipofectamine 2000 (Sigma). For electrophysiology experiments, cells were transfected with wild-type-mGluR2 or wild-type-mGluR3, GIRK1-F137S, and yellow fluorescent protein (YFP) (as a transfection marker) at a 7:7:1 ratio with 0.7 μg plasmid per well for receptor and channel. For FRET experiments, cells were transfected with SNAP and CLIP-tagged constructs at a ratio of 1:2 with 0.3 μg of SNAP–mGluR DNA per well.

Patch clamp electrophysiology. Whole-cell patch clamp recordings from single isolated cells were performed 24–48 h after transfection in a high potassium extracellular solution containing (in mM): 120 KCl, 29 NaCl, 1 MgCl₂, 2 CaCl₂ and 10 HEPES, pH 7.4. Cells were voltage clamped to −60 mV using an Axopatch 200B amplifier (Axon Instruments) and membrane currents were recorded. Glass pipettes of resistance between 3 and 8 MΩ were filled with intracellular solution containing (in mM): 140 KCl, 10 HEPES, 3 Na₂ATP, 0.2 Na₂GTP, 5 EGTA and 3 MgCl₂, pH 7.4. Data were acquired with a 2 kHz acquisition rate and filtered with the amplifier 4-pole Bessel filter at 1 kHz. Data acquisition and analysis were performed using pCLAMP 10 software (Axon Instruments).

Ensemble FRET. Approximately 24–48 h after transfection, cells were labelled while attached to poly-L-lysine-coated coverslips. Culture media was removed and coverslips were washed and transferred to extracellular solution containing (in mM): 135 NaCl, 5.4 KCl, 2 CaCl₂, 1 MgCl₂, 10 HEPES, pH 7.4. Cells were labelled at 37 °C with 2.5 μM benzylguanine Alexa-647 (NEB) for 45 min followed by 5 μM benzylcytosine DY-547 (NEB) for 45 min. The fluorophores were diluted in extracellular solution and coverslips were washed in between labelling with donor and acceptor. After labelling, cells were mounted on an upright, scanning confocal microscope (Zeiss LSM 780) and imaged with a ×20 objective. Donor excitation was performed using a 561-nm laser and images were taken in the donor and acceptor channels at 1 Hz. Clusters of cells were analysed together and FRET was calculated as FRET = $(I_A)/(I_D + I_A)$, in which I_D is the fluorescence donor intensity, and I_A is the fluorescence acceptor intensity. For individual traces, FRET was normalized to the basal FRET value observed before application of any drugs. FRET changes calculated for dose–response curves were normalized to saturating glutamate (1 mM) and dose–response curves were obtained from multiple cell clusters and averaged from at least three experiments. Fitting of dose–response curves was performed using Prism (Graphpad). All drugs were purchased from Tocris and delivered with a gravity-driven perfusion system.

smFRET measurements. To inhibit nonspecific protein adsorption, flow cells for single-molecule experiments were prepared as previously described using mPEG (Laysan Bio) passivated glass coverslips (VWR) and doped with biotin PEG[16]. Before each experiment, coverslips were incubated with NeutrAvidin (Thermo), followed by 10 nM biotinylated secondary antibody (donkey anti-rabbit, Jackson ImmunoResearch). For receptor immunopurification, 10 nM anti-mGluR2 primary antibody (Cell Signaling, 12056) or 10 nM anti-mGluR3 antibody (abcam), or 15 nM anti-HA antibody (abcam, ab26228) was incubated in the chamber (Fig. 1c). Between each conjugation step, the chambers were flushed to remove free reagents. The antibody dilutions and washes were done in T50 buffer (50 mM NaCl, 10 mM Tris, pH 7.5).

For single-molecule experiments, fresh cells expressing tagged mGluR constructs were labelled as described above. After labelling, cells were recovered from coverslips by incubating with Ca²⁺-free PBS buffer for 20–30 min followed by gentle pipetting. Cells were then pelleted and lysed in the lysis buffer consisting of 150 mM NaCl, 1 mM EDTA, protease inhibitor cocktail (Thermo Scientific) and 1.2% IGEPAL (Sigma) or 1% n-dodecyl-B-D-maltoside (DDM) (Anatrace). After 1 h incubation at 4 °C, cells were centrifuged at 16,000g for 20 min and supernatant was collected and kept on ice. The cell lysate was diluted to achieve sparse immobilization of labelled receptors on the surface (ranging from 5× to 50× dilution depending on the expression and labelling efficiency) and applied to coverslips. After achieving optimum surface immobilization (~400 molecules in a 2,000 μm²

imaging area), unbound receptors were washed out of the flow chamber and the flow cells were then washed extensively (up to 50× the cell volume). Finally, receptors were imaged in the imaging buffer consisting of (in mM) 3 Trolox, 120 KCl, 29 NaCl, 2 CaCl₂, 1 MgCl₂, 50 HEPES, 0.04% IGEPAL and an oxygen scavenging system (0.8% dextrose, 0.8 mg ml⁻¹ glucose oxidase, and 0.02 mg ml⁻¹ catalase), pH 7.4. Reagents were purchased from Sigma and were all UltraPure grade (purity >99.99%). All buffers were made in UltraPure distilled water (Invitrogen). For the experiments done in the absence of Ca²⁺, 10 mM EGTA and 1 mM EDTA were added to the imaging buffer. Catalase was diluted in T50 buffer and passed through a spin column 3× (BioRad). To ensure lack of glutamate contamination for the experiments done in the absence of glutamate further, the oxygen scavenging solution was treated with glutamic-pyruvic transaminase (Sigma) in the presence of 2 mM sodium pyruvate (Gibco) to remove possible trace levels of glutamate.

Samples were imaged with a ×60 objective (Olympus) on a total internal reflection fluorescence microscope with 30 ms time resolution unless stated otherwise. Lasers 532 nm (Cobolt) and 632 nm (Melles Griot) were used for donor or acceptor excitation, respectively. FRET efficiency was calculated as $(I_A - 0.1I_D)/(I_D + I_A)$, in which I_D and I_A are the donor and acceptor intensity after background subtraction.

smFRET data analysis. Single-molecule intensity traces showing single-donor and single-acceptor photobleaching with a stable total intensity for longer than 5 s were collected (20–30% of total molecules per imaging area). Individual traces were smoothed using a nonlinear filter from ref. 30 with filter parameters: window = 2, $M = 2$ and $P = 15$. Each experiment was repeated at least five times independently to ensure reproducibility of the results and one data set for each condition is presented. FRET histograms were compiled from at least 200 molecules per condition (30 ms time resolution). Error bars in the histograms represent the standard error from at least six independent movies. To ensure that traces with different length contribute equally, histograms from individual traces were normalized to one before compiling. Fitting to histograms was done in Origin Pro. Synchronized density plots were constructed from manually selected transitions that were synchronized to the data point where FRET signal drops below (forward transition) or above (reverse transition) the threshold of 0.3. The proportion of dynamic traces was calculated manually from the raw smFRET traces. Each molecule that exhibited at least one FRET transition (defined by anti-correlated changes in donor and acceptor signals) and lasted for at least two data points was counted as having dynamics. Dwell-time analysis was performed using unfiltered traces to avoid smoothing artefacts. First, donor and acceptor signals were each idealized to a two state signal with the value of idealized intensity obtained from the two main peaks in the intensity histogram. Idealized FRET was then calculated from idealized donor and acceptor signals and further analysed to calculate the active state dwell time. Owing to time resolution limitations, dwell-time measurement represents an upper limit estimate. All transitions were individually inspected to make sure the FRET transitions were real based on anti-correlation of the donor and acceptor intensity signals at the transition. The number of traces used in dwell-time analysis for mGlu2 with glutamate, LY379268 and DCG-IV are 53, 61 and 41 obtained from at least six independent movies and error bars represent the s.e.m. More than 300 transitions in each case were analysed. The hidden Markov fit in Fig. 2 was done using HaMMy[31]. The cross-correlation (CC)[32] of donor and acceptor intensity time traces at time τ is defined as $CC(\tau) = \delta I_D(t)\delta I_A(t + \tau)/(<I_D(t)> + <I_A(t)>)$, in which $\delta I_D(t) = I_D(t) - <I_D(t)>$, and $\delta I_A(t) = I_A(t) <I_A(t)>$. $<I_A(t)>$ and $<I_D(t)>$ are time average donor and acceptor intensities, respectively. The cross-correlation calculation was performed on the same traces that were used for the histogram. We fit the cross-correlation data to a single exponential function to obtain two parameters: the characteristic time of the exponential (τ) and the amplitude of the exponential.

Because of the low anisotropy of the donor and acceptor fluorophores when attached to these SNAP and CLIP tags[24], and taking into account their spectral overlap, a Förster radius of 52 Å was used to estimate the distances.

30. Haran, G. Noise reduction in single-molecule fluorescence trajectories of folding proteins. *Chem. Phys.* **307,** 137–145 (2004).
31. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91,** 1941–1951 (2006).
32. Zhou, R. *et al.* SSB functions as a sliding platform that migrates on DNA via reptation. *Cell* **146,** 222–232 (2011).
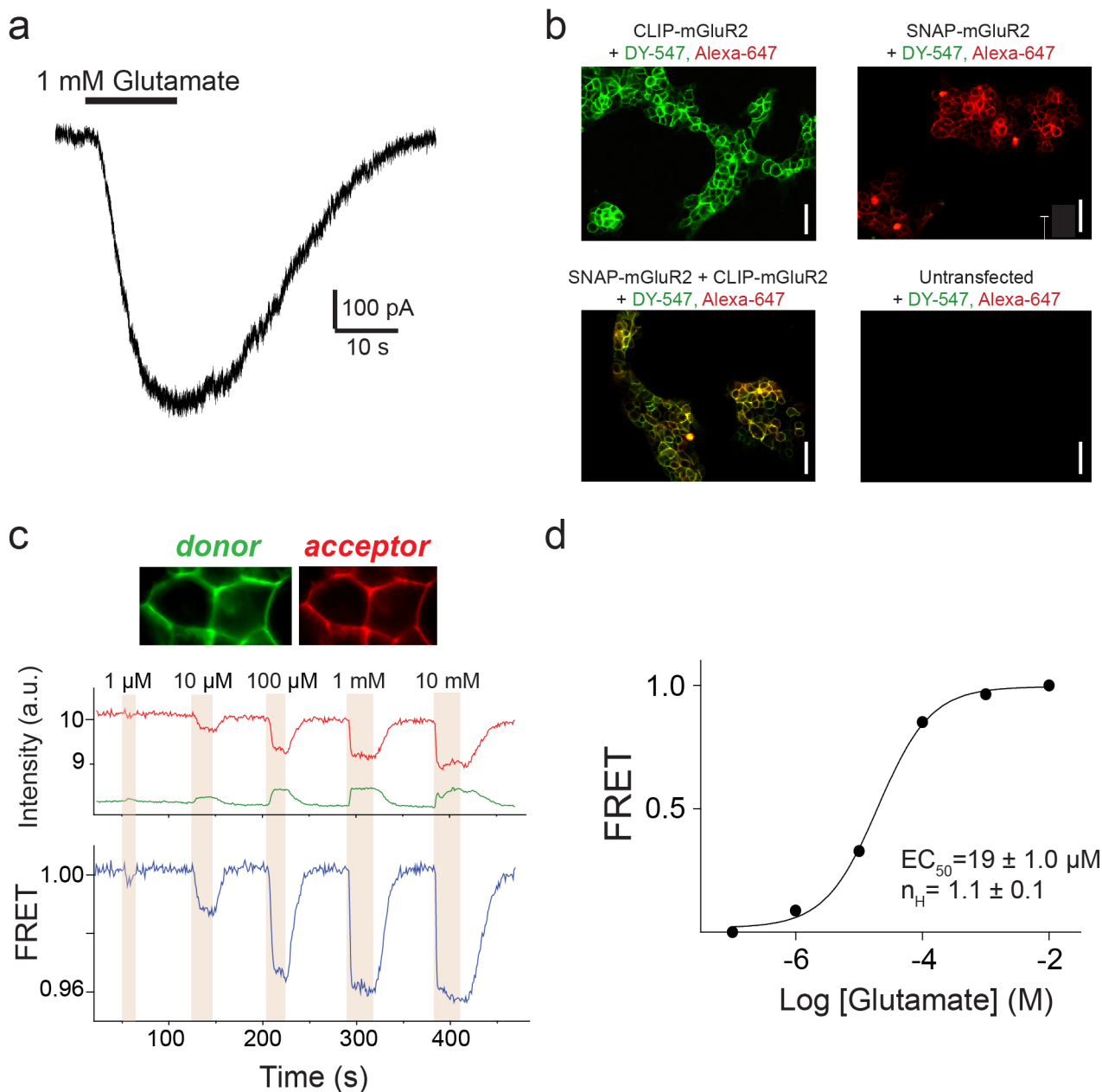
Extended Data Figure 1 | Fluorophore labelling is specific and FRET constructs are functional in HEK293T cells. a, SNAP–mGluR2 shows glutamate-induced currents in cells co-expressing GIRK channels. **b**, Treatment with 2.5 µM benzylguanine Alexa-647 (for SNAP) followed by 5 µM benzylcytosine DY-547 (for CLIP) produces specific and orthogonal labelling of SNAP and CLIP–mGluR2 constructs in HEK293T cells. All conditions were imaged with identical settings in both the red (excitation = 635 nm) and green (excitation = 561 nm) channels. **c**, Ensemble FRET measurements from HEK293T cells. Top, image of cells expressing SNAP and CLIP–mGluR2 labelled with donor (DY-547, green) and acceptor (Alexa-647, red). Bottom, representative trace showing dose-dependent, reversible decrease in FRET after glutamate application. Glutamate was washed out between applications. **d**, Ensemble FRET glutamate titration in HEK293T cells. Error bars are s.e.m.
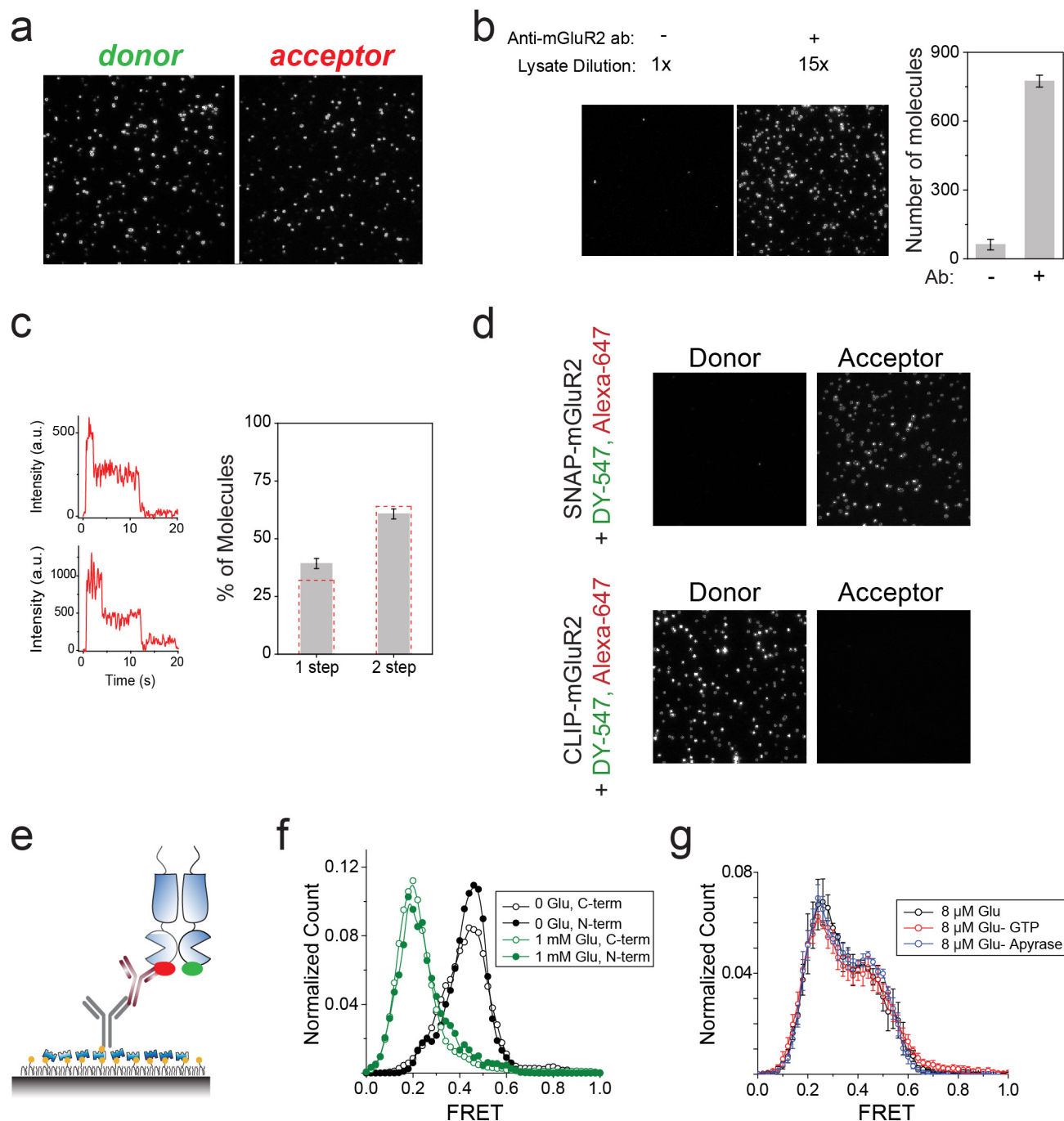
Extended Data Figure 2 | Control experiments verifying specificity of the smFRET assay. a, Representative TIRF image showing single receptors in donor and acceptor channels during donor excitation with a 532 nm laser. **b,** Single-molecule pull-down of SNAP–mGluR2 on a passivated surface is specific. Left, representative images of individual molecules in the absence or presence of an anti-mGluR2 antibody. Right, quantification of the number of molecules pulled down for each condition. **c,** Photobleaching step analysis shows that mGluR2 remains a dimer in single-molecule pull-down. Left, representative single-molecule bleaching steps for mGluR2–GFP. Right, histogram of bleaching step counts for all molecules. Dotted red line shows the predicted proportions for an 80% GFP maturation rate. **d,** Pulldown of lysate from cells expressing only SNAP–mGluR2 (**c**) or CLIP–mGluR2 (**d**), and labelled with both donor and acceptor fluorophores confirms labelling specificity at the single-molecule level. **e, f,** Pulldown of SNAP– and CLIP–mGluR2 via an antibody against an N-terminal HA-tag (**e**) leads to very similar smFRET histograms (**f**, filled circles) in the absence (black) or presence of 1 mM glutamate (green) compared to pull-down with a C-terminal antibody (**f**, open circles). **g,** Application of either GTP, to remove any co-assembled G proteins, or apyrase, to lock any G proteins onto mGluR2, does not alter smFRET histograms.
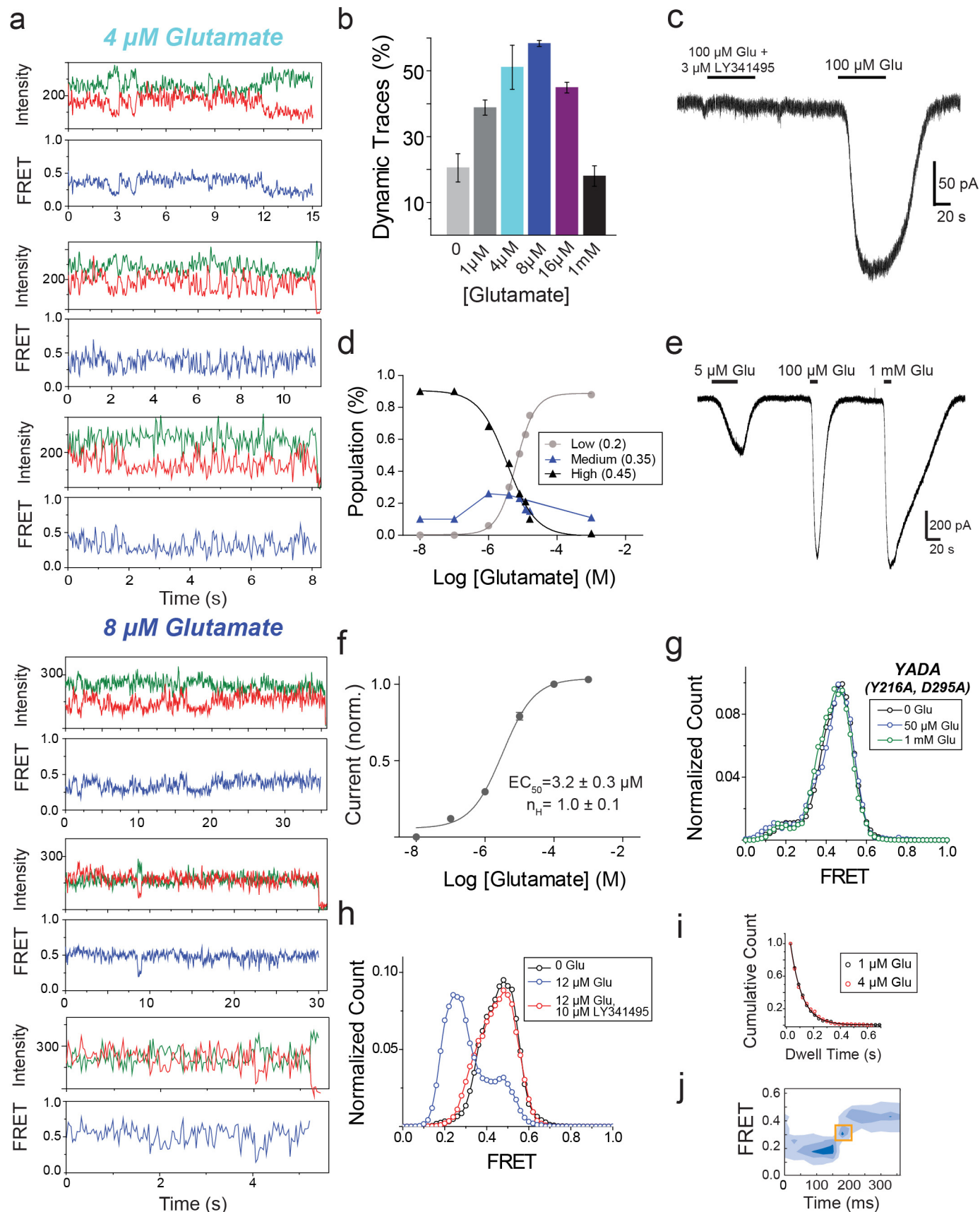
Extended Data Figure 3 | Further analysis of glutamate-induced smFRET and functional properties of mGluR2. **a**, Representative smFRET traces for mGluR2 in the presence of 4 µM or 8 µM glutamate. **b**, Quantification of the percentage of single-molecule traces showing at least one transition to the active state at different glutamate concentrations **c**, In HEK293T cells co-expressing mGluR2 and GIRK, LY341495 prevents glutamate-induced inward currents without altering the baseline current. **d**, Glutamate titration curves produced from fitting FRET histograms to the sum of three Gaussian distributions. **e, f**, Glutamate induces inward currents via mGluR2 in a dose-dependent manner ($n = 9$ cells). **g**, The glutamate-insensitive mGluR2-YADA (Tyr216Ala, Asp295Ala) shows no smFRET response to 50 µM or 1 mM glutamate. **h**, smFRET histograms showing that application of the competitive antagonist LY341495 reverses the FRET change induced by glutamate. **i**, Dwell time analysis of mGluR2 for sub-saturating glutamate concentrations. Solid lines show single exponential fits to the data. **j**, FRET density plots constructed from synchronized transitions from the low to high FRET states show a short dwell at the medium FRET level (yellow box). Error bars are s.e.m.
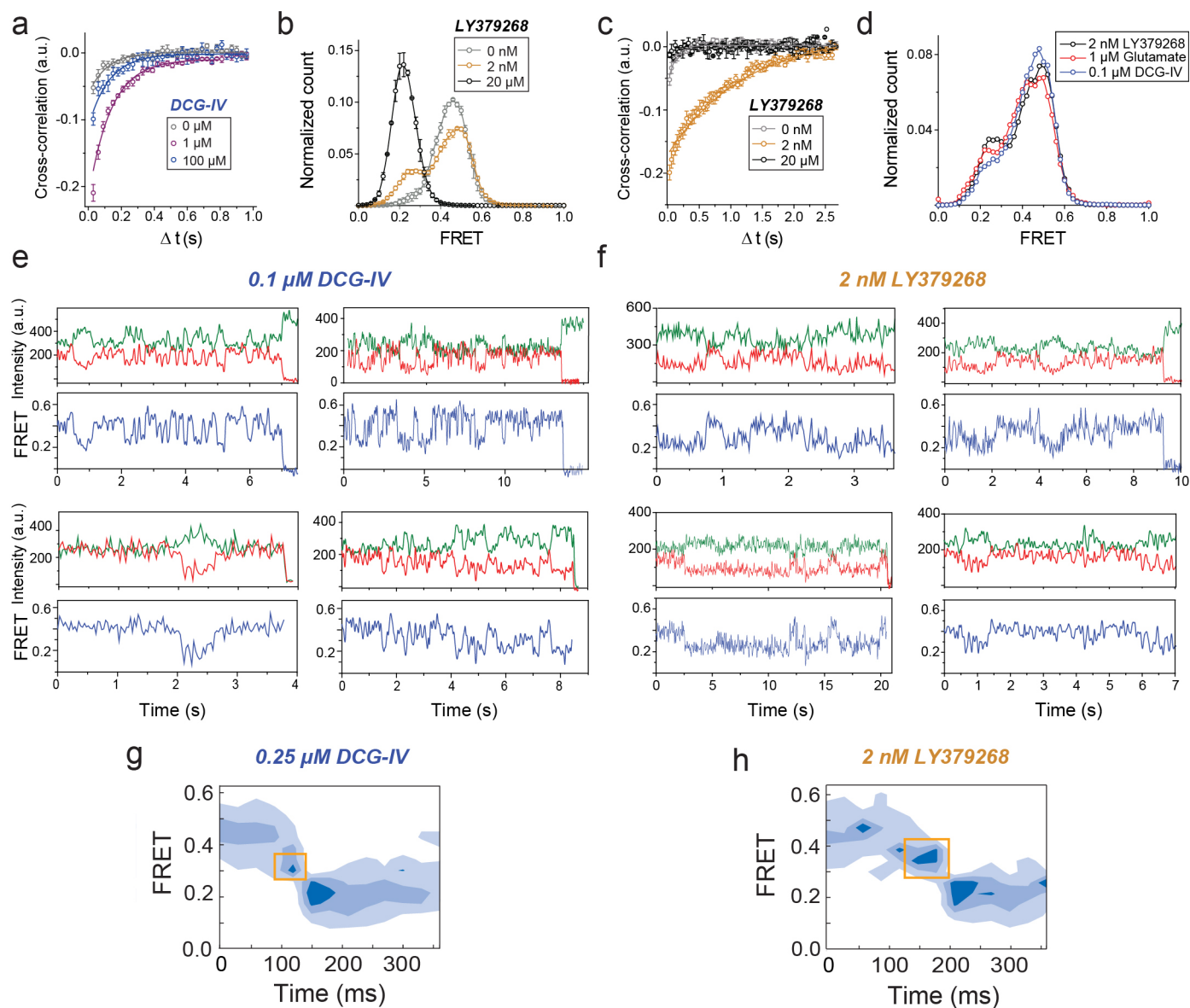
Extended Data Figure 4 | Further analysis of the effects of orthosteric agonists on mGluR2 smFRET. a, Cross-correlation plot for mGluR2 in the presence of DCG-IV shows concentration-dependent dynamics. **b, c**, smFRET histogram (**b**) and cross-correlation plots (**c**) for mGluR2 in the presence of the full agonist LY379268. **d**, smFRET histograms in the presence of 1 μM glutamate, 100 nM DCG-IV and 2 nM LY379268 yields comparable occupancy of the active state. **e, f**, Representative smFRET traces for mGluR2 in the presence of 0.1 μM DCG-IV (**e**) or 2 nM LY379268 (**f**), which are the concentrations used for dwell-time analysis. **g, h**, FRET density plots constructed from the average transitions from the low to high states show a short dwell at the medium FRET level (yellow boxes) for mGluR2 in the presence of DCG-IV (**g**) or LY379268 (**h**).
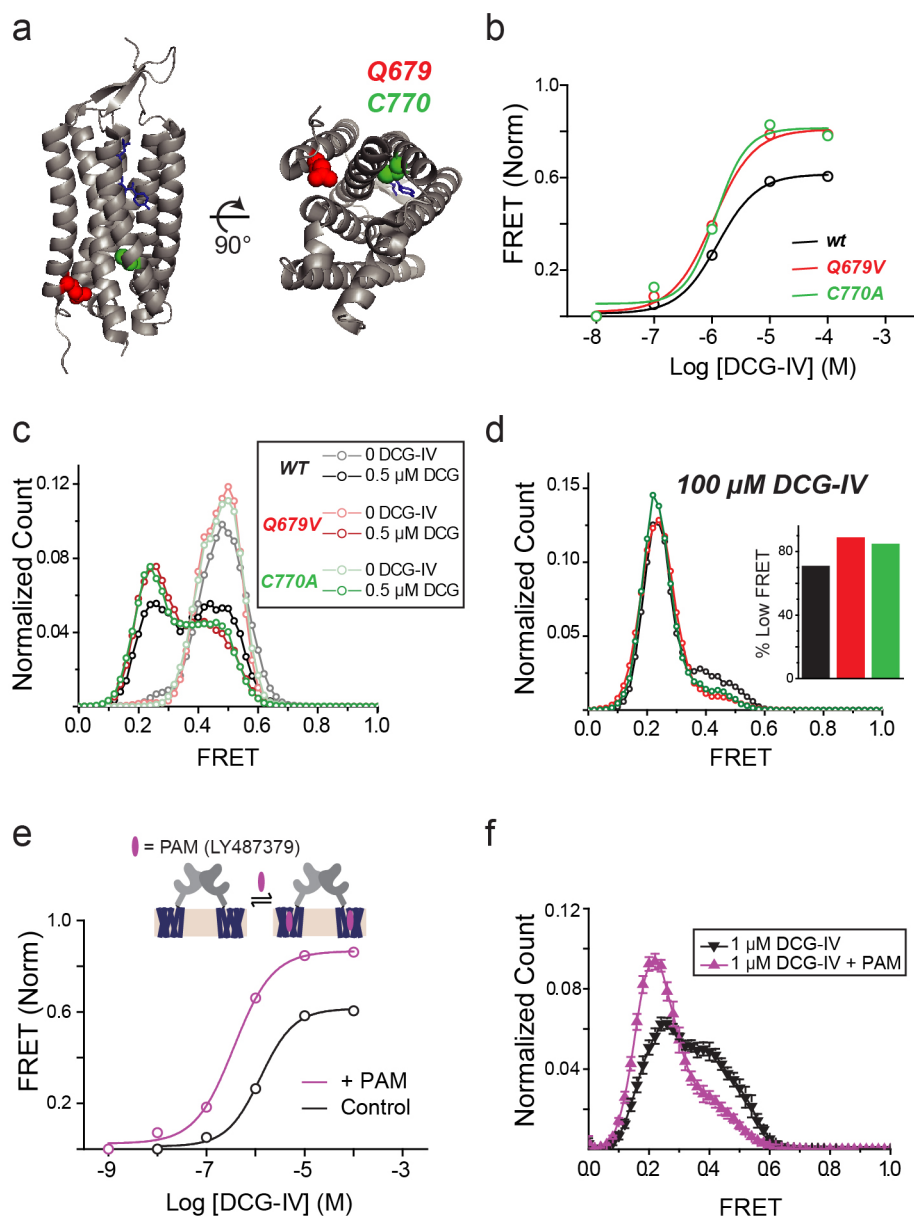
Extended Data Figure 5 | TMD mutations that introduce basal activity and a positive allosteric modulator increase the affinity and efficacy of a partial agonist. a, Crystal structures of the mGluR1 TMD bound to a negative allosteric modulator (PDB code 4OR2) showing the location of conserved residues in TM4 and TM6 previously shown to be sensitive to mutations that induce basal activity. **b**, Ensemble FRET titrations showing that mutations Gln679Val and Cys770Ala increase the affinity and efficacy of DCG-IV compared to wild-type mGluR2. **c, d**, smFRET histograms for TMD mutants show population of the same three FRET states as the wild type, but with greater occupation of the low FRET state at either sub-saturating (**c**) or saturating (**d**) concentrations of DCG-IV. **e**, Binding of PAM LY487379 to the TMD of mGluR2 (top) increases the apparent affinity and efficacy of DCG-IV in ensemble FRET measurements in HEK293T cells. All values were normalized to the response to 1 mM glutamate. **b**, smFRET histograms showing a LY487379 -induced shift in the response to 1 μM DCG-IV. Error bars are s.e.m.
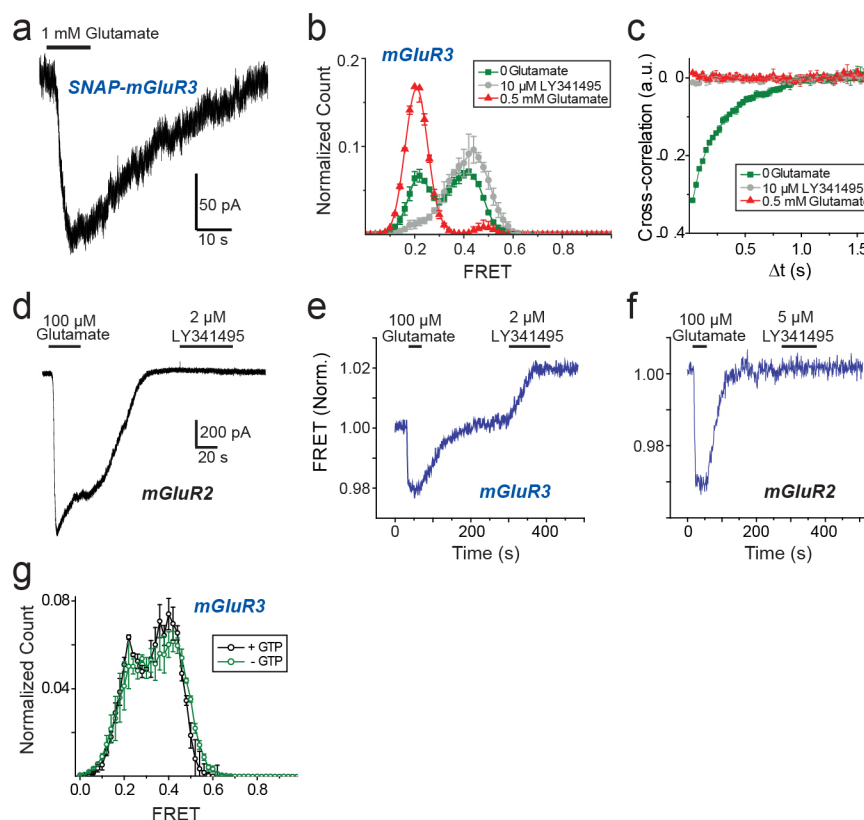
Extended Data Figure 6 | Characterization of basic ensemble and smFRET properties of mGluR3 as compared to mGluR2. **a**, Activation of GIRK by SNAP–mGluR3 in HEK293T cells. **b**, smFRET histograms for mGluR3 show glutamate-independent low FRET population. **c**, Cross-correlation plots for mGluR3 show glutamate-independent dynamics. **d**, Unlike mGluR3, mGluR2 shows zero or minimal current response to the antagonist LY341495 in the absence of glutamate. **e**, **f**, Ensemble FRET in HEK293T cells shows a robust antagonist LY341495-induced FRET increase in mGluR3 (**e**) but not in mGluR2 (**f**). **g**, smFRET histograms for mGluR3 with or without GTP treatment to dissociate any G proteins that may be coupled to the receptor. The time resolution for this data is 100 ms. Error bars are s.e.m.
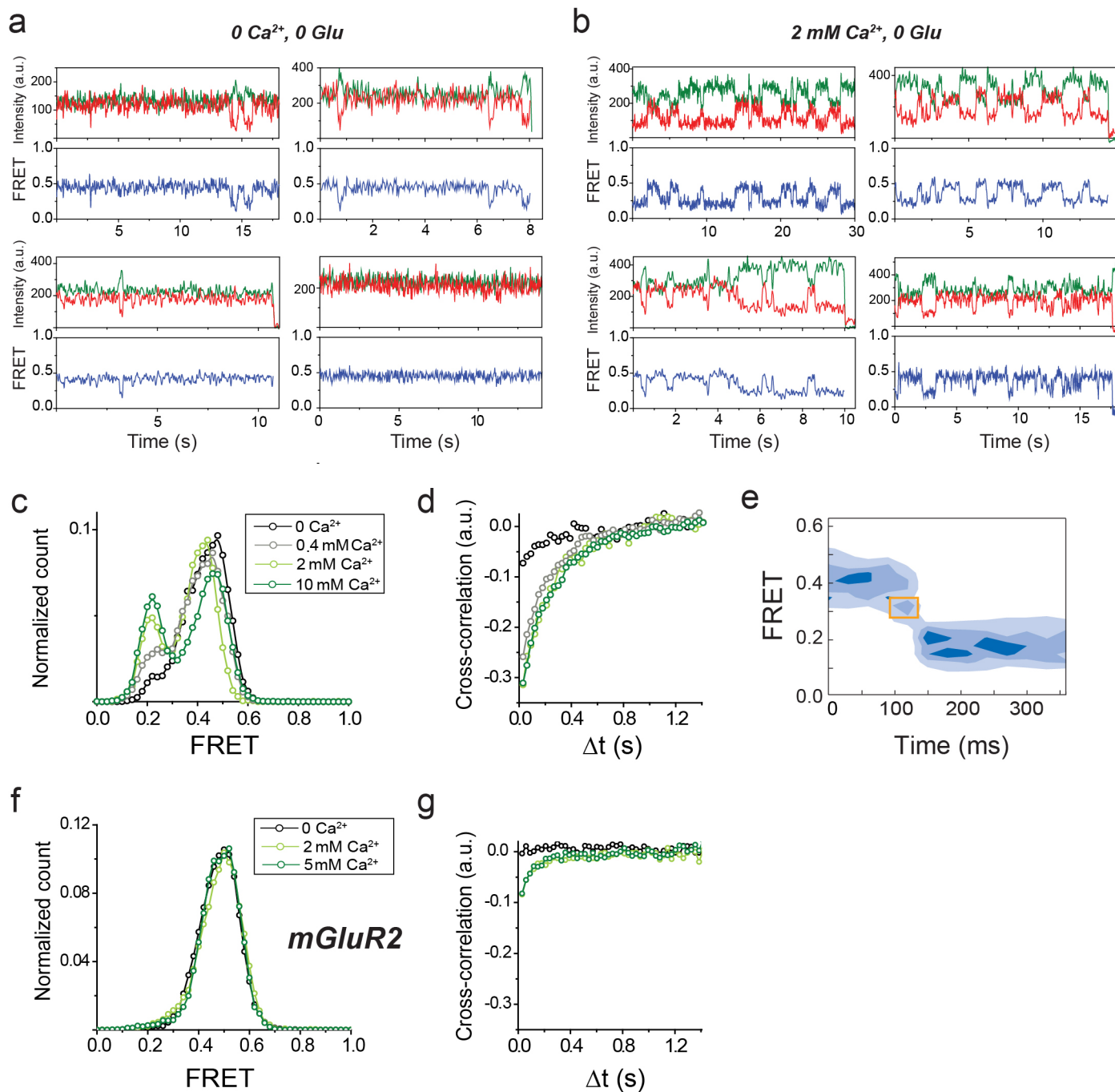
Extended Data Figure 7 | Calcium sensitivity of mGluR3. a, b, Representative smFRET traces for mGluR3 in the absence (**a**) or presence (**b**) of 2 mM Ca^{2+}. **c,** smFRET histograms for mGluR3 in the presence of various concentrations of calcium. **d,** Cross-correlation plots for mGluR3 in the presence of various concentrations of calcium. **e,** FRET density plot for showing that Ca^{2+}-induced synchronized transitions show a similar intermediate in mGluR3, as observed in mGluR2. **f, g,** smFRET histograms (**f**) and cross-correlation plots (**g**) for mGluR2 in the presence of various concentrations of calcium.
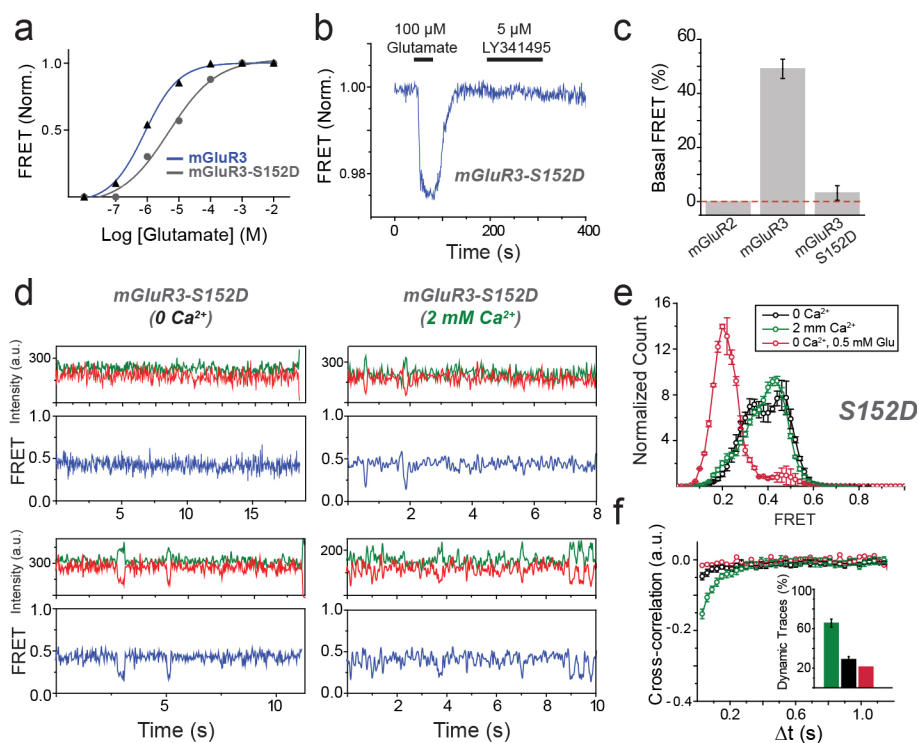
Extended Data Figure 8 | mGluR3(Ser152Asp) shows decreased basal FRET and calcium sensitivity. a, Ensemble FRET glutamate titrations for mGluR3 and mGluR3(Ser152Asp). **b**, Representative ensemble FRET trace for mGluR3(Ser152Asp) shows no response to LY341495. **c**, Summary of basal FRET for mGluR3, mGluR2 and mGluR3-S152D. Basal FRET = [ΔFRET$_{LY341495}$]/([ΔFRET$_{LY341495}$] + [ΔFRET$_{Glu}$]). **d**, Representative smFRET traces for mGluR3(Ser152Asp) in the absence (left) or presence (right) of 2 mM Ca^{2+}. **e**, smFRET histograms for mGluR3(Ser152Asp) in the absence or presence of Ca^{2+} or saturating glutamate. **f**, Cross-correlation plots for mGluR3(Ser152Asp). Inset shows the percentage of traces showing dynamics in different ligand conditions. Error bars are s.e.m.
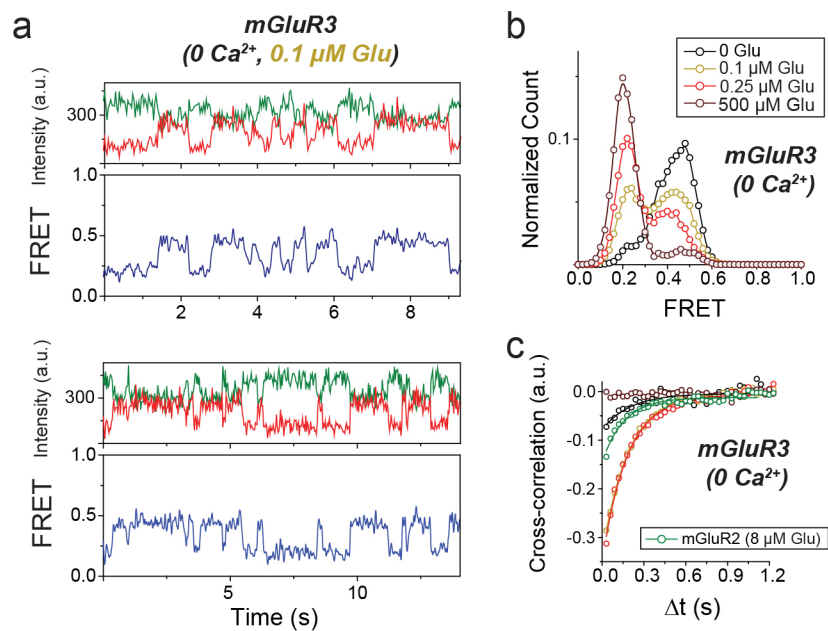
Extended Data Figure 9 | Glutamate induced smFRET dynamics of mGluR3 in the absence of calcium. a, Representative smFRET traces for mGluR3 in the absence of Ca^{2+} and the presence of sub-saturating glutamate. **b,** smFRET histogram showing dose-dependent response of mGluR3 to glutamate in the absence of Ca^{2+}. **c,** Cross-correlation plots showing glutamate-induced smFRET dynamics. mGluR2 (green) at its maximum dynamics shows a smaller cross-correlation amplitude compared to mGluR3.
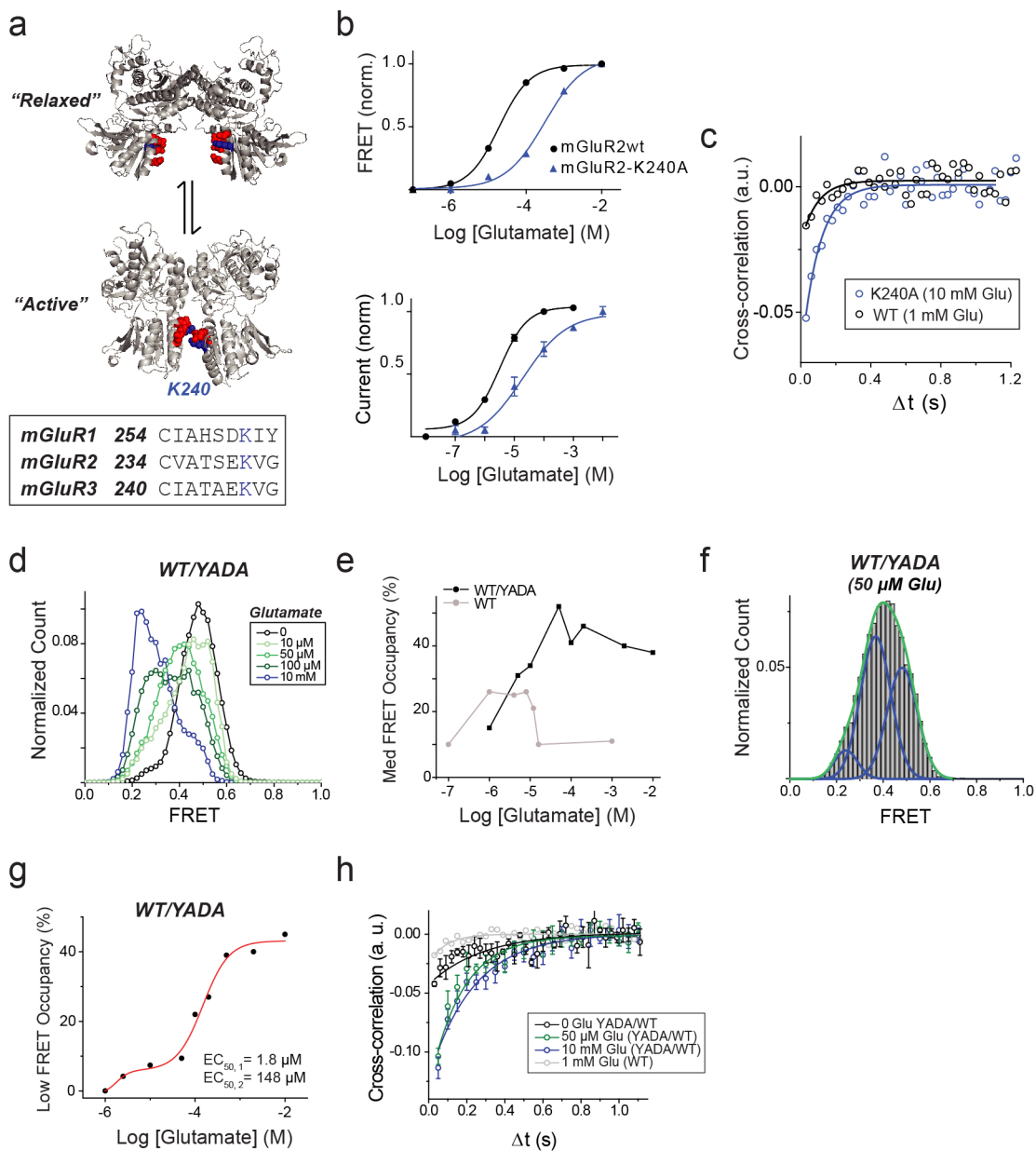
Extended Data Figure 10 | Further characterization of mGluR2(Lys240Ala) and wild-type/YADA heterodimers. a, Crystal structures of mGluR1 in the relaxed (top; PDB code 1EWT) and active states (bottom; PDB code IEWK) show a reorientation of the dimer interface that brings charged residues of the lower lobe in close proximity. Conserved negatively charged residues are shown in red, and Lys260 (Lys240 in mGluR2) is shown in blue. **b**, Glutamate titrations show a decreased apparent affinity for mGluR2(Lys240Ala) in ensemble FRET (top) and GIRK current activation (bottom). **c**, Cross-correlation plots in the presence of saturating glutamate show increased dynamics for mGluR2(Lys240Ala) relative to wild type. **d**, smFRET histogram showing distributions for wild-type/YADA heterodimers at a range of glutamate concentrations. **e**, Concentration-dependence of low FRET population in wild-type/YADA heterodimers produced from fitting FRET histograms to the sum of three Gaussian distributions. The EC_{50} for each phase of the distribution is shown. **g**, Concentration-dependence of medium FRET population in wild-type/YADA heterodimers and wild-type homodimers produced from fitting FRET histograms to the sum of three Gaussian distributions. **g**, Three-state fit to FRET histogram for wild-type/YADA heterodimers in the presence of 100 μM glutamate shows substantial population of the medium FRET (0.35) peak. **h**, Cross-correlation plots for wild-type/YADA heterodimers. Error bars are s.e.m.

CORRECTIONS & AMENDMENTS

Corrigendum: *In vivo* engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system

Danilo Maddalo, Eusebio Manchado, Carla P. Concepcion, Ciro Bonetti, Joana A. Vidigal, Yoon-Chi Han, Paul Ogrodowski, Alessandra Crippa, Natasha Rekhtman, Elisa de Stanchina, Scott W. Lowe & Andrea Ventura

During the assembly of Fig. 1b of this Letter, the sequences of the targeted regions in *Eml4* and *Alk*, shown in the schematic in the top panel, were inadvertently swapped. In addition, the legend for the top panel of Fig. 1b should have read: "**b**, A schematic of the loci before and after the inversion with the location of the primers used (top panel). The sequences recognised by the guide RNAs are shown in italics, the protospacer adjacent motif (PAM) sequences are in bold and underlined. Asterisks indicate the predicted Cas9 cleavage sites (positions −3 and −4 from the PAM).". The corrected panel is shown in Fig. 1 of this Corrigendum. We apologize for the confusion this may have caused.
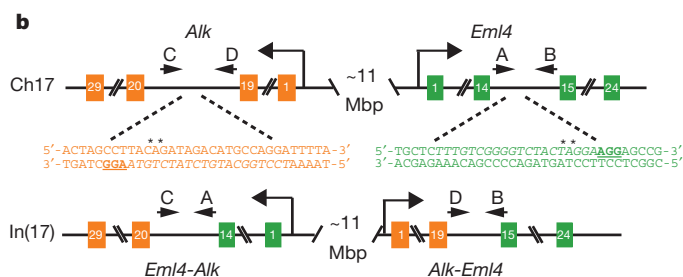
Figure 1 | This is the corrected top panel of Fig. 1b of the original Letter.

CORRECTIONS & AMENDMENTS

Erratum: No signature of ejecta interaction with a stellar companion in three type Ia supernovae

Rob P. Olling, Richard Mushotzky, Edward J. Shaya, Armin Rest, Peter M. Garnavich, Brad E. Tucker, Daniel Kasen, Steve Margheim & Alexei V. Filippenko

Owing to a typesetter error, the final word 'days' was missing from the end of the Table 1 footnote in the PDF and print versions (the HTML is correct). The sentence has been corrected in the PDF version to read "All listed times are with respect to maximum light, in rest-frame days."

CORRECTIONS & AMENDMENTS

Erratum: A strong ultraviolet pulse from a newborn type Ia supernova

Yi Cao, S. R. Kulkarni, D. Andrew Howell, Avishay Gal-Yam,
Mansi M. Kasliwal, Stefano Valenti, J. Johansson, R. Amanullah,
A. Goobar, J. Sollerman, F. Taddia, Assaf Horesh, Ilan Sagiv,
S. Bradley Cenko, Peter E. Nugent, Iair Arcavi, Jason Surace,
P. R. Woźniak, Daniela I. Moody, Umaa D. Rebbapragada,
Brian D. Bue & Neil Gehrels

In this Letter, the superscript in the ultraviolet luminosity was listed incorrectly as '-41' rather than '41' in the last sentence of the second paragraph from the bottom in the left column of page 1. It should have read $L_{UV} \approx 3 \times 10^{41}$ erg s^{-1}. This has been corrected online.

THE DNA OF A NATION

The United Kingdom aims to sequence 100,000 human genomes by 2017. But screening them for disease-causing variants will require innovative software.

BY VIVIEN MARX

Soon after the Californian twins were born, their parents grew concerned: the children were developing slowly and had floppy muscle tone. A brain scan indicated that the boy might have cerebral palsy, but doctors were puzzled over his sister's tremor and seizures. Batteries of tests failed to confirm diagnoses in either child, and treatment when the children were five with the drug L-dopa — used for people with Parkinson's disease — helped only for a while.

It was only in 2010, when the twins reached the age of 14, that whole-genome sequencing ended their diagnostic odyssey. It identified a pair of mutations in a gene that encodes the enzyme sepiapterin reductase, which is involved in production of the neurotransmitters dopamine and serotonin. Doctors modified the treatment to include serotonin; the boy's mobility improved, and the girl was no longer plagued by sudden, breath-stealing spasms[1].

Stories such as this one fuel ambitions to diagnose more quickly and accurately using genomic medicine. Indeed, tests that can probe certain disease-associated genes are increasingly becoming a diagnostic option.

But such genetic tests often fail to give a diagnosis because they are too focused on a selection of known genes on one section of the genome. In cases like that of the twins, researchers or clinicians must go further and sample a person's whole genetic sequence to find the disease-causing genes. Currently this is done only in rare cases — but a number of large-scale initiatives are poised to bring whole-genome analysis into routine medical care.

The United Kingdom has taken a giant leap into genomic medicine with the 100,000 Genomes Project, which was launched in 2012 and has been personally backed by Prime Minister David Cameron. As part of the £300-million (US$467-million) initiative, 100,000 genomes from National Health Service (NHS) patients with cancer, rare disorders and infectious diseases will be sequenced by 2017. The project's aims are to gain scientific insight by linking the disorders with precise genetic signatures; to obtain better diagnoses; to tailor treatments to individual patients; and, ultimately, to spur ▶

▶ the development of a UK genomics industry.

The state-funded, centralized UK healthcare system is ideal for such population-based approaches in genomic medicine, says John Bell, who is a medical researcher at the University of Oxford, UK, and is also on the board of Genomics England, the NHS-owned company set up to run the project. The NHS already holds extensive clinical information on individuals, and pairing this with detailed genomic data will enable powerful insights into the links between medicine and genetics. Evidence that whole-genome interpretation can help in a wide range of disorders is mounting[2], and in the long term, Bell says, the goal is to make whole genomes part of regular NHS health records.

But before that vision can be realized, there are several hurdles that the 100,000 Genomes Project must overcome. Aside from the logistical task of extracting and sequencing DNA from thousands of individuals, there is the problem of identifying which genome variations cause disease and which are harmless — a daunting, data-heavy and time-consuming process that will require a slew of specialized companies with dedicated software.

CONSIDERABLE COHORT

Iceland was the first to launch a large-scale genomic analysis of its population. Many nations have followed suit with the explicit goal of linking health care and genomics. In the United States, the Precision Medicine Initiative plans to sequence the genomes of one million volunteers, and the Million Veteran Program is gearing up to do likewise with US military veterans. Similar projects are under way in Canada, Australia, Japan, South Korea, Singapore, Thailand, Kuwait, Qatar, Israel, Belgium, Luxembourg and Estonia.

But the 100,000 Genomes Project is the venture gaining the most steam: it has already enrolled 3,500 people with rare diseases and 2,000 individuals with cancer, and will involve roughly 75,000 people altogether (see 'The clinical genome'). People with rare diseases and their relatives will make up 50,000 of the final figure; 80% of rare diseases are inherited, so the genome of the affected person (usually a child) will be sequenced along with the genomes of two of their closest blood relatives. The remaining group of 25,000 will be composed of people with cancer, who will have their genome sequenced twice (the tumour DNA will be compared with that from a patient's normal cells), giving the grand total of 100,000 genome sequences.

The hope is that participants will benefit from clinical insights into their condition. But their genomes will also contribute knowledge of value to the entire patient community. One person's prostate-cancer genome, for example, might reveal specific genetic patterns that a physician can compare against the Genomics England database. The physician can then find other people with similar patterns and learn which drugs and procedures worked best for them.

INDUSTRY PARTNERSHIP

Genomics England is now selecting industry partners for each step in the process, from extracting the DNA to interpreting the genome. Sequencing-instrument manufacturer Illumina, which has its headquarters in San Diego, California, is handling the sequencing as well as the job of identifying genetic variants — known as variant calling. This is being done from Illumina's site in Little Chesterford, UK, but the company will be moving its sequencing instruments to the Wellcome Trust Genome Campus in Hinxton, UK, over the next few months as the project starts to scale up.

Illumina is processing extracted DNA using high-throughput sequencing to obtain short fragments of the strings of As, Ts, Cs and Gs that are the building blocks of DNA. These fragments are computationally assembled back into a contiguous sequence and then, using bioinformatics, scientists will compare the resulting complete genomes with the human reference genome: a representative example of the human genome that is continually updated by the international Genome Reference Consortium. The aim is to note each deviation from this reference: each genetic variant.

To identify these variants, the Illumina team will use the company's Isaac workflow, an open-source computational genome-alignment and variant-calling tool[3]. Then, to discover which of each genome's hundreds of thousands of variants have a role in the individual's disease, Genomics England will gather and analyse the sequences and variants from all 100,000 genomes at its secure data centre in Corsham, UK. But the NHS company has still to decide which software to use for this process: it will work with Illumina and academics to test, benchmark and improve the many existing variant-calling algorithms before selecting the final software for the initiative.

> *"The goal is to make whole genomes part of regular NHS health records."*

Illumina has already sequenced more than 3,000 genomes for the project, and more are pouring in daily. For the variant analysis, the firm will collect and report different types of germline and somatic variants found at specific locations in DNA, says Peter Fromen, the company's managing director of population-sequencing initiatives. Variants can include insertions or deletions of a few nucleotides, or the substitution of one nucleotide for another. There can also be structural variants, such as changes in the copy number of a gene.

Each variant is then checked against those in existing variant databases, such as dbSNP, a short-genetic-variation database curated at the US Natio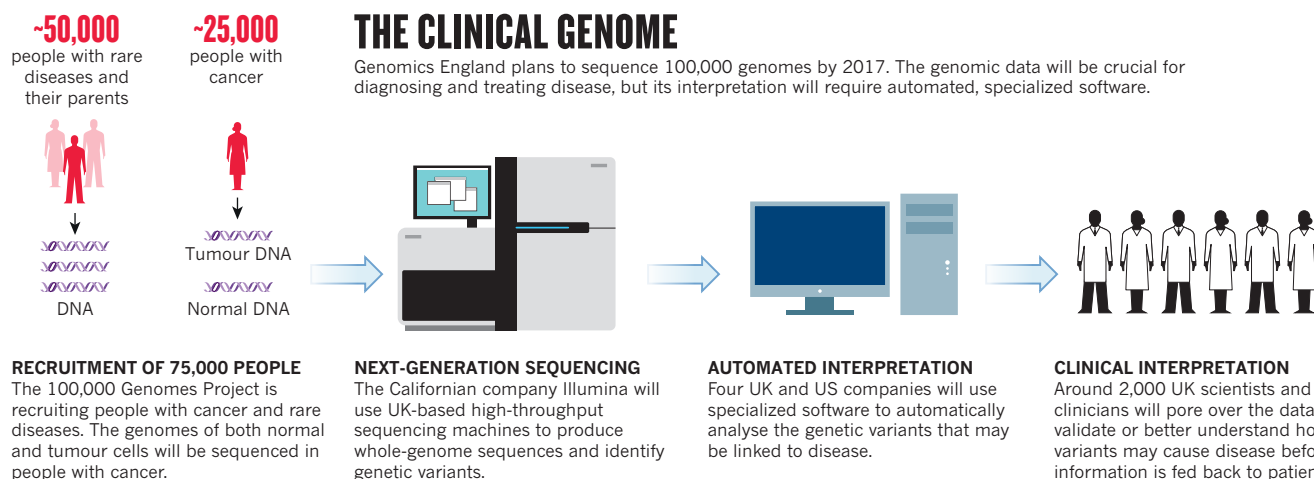nal Institutes of Health (NIH); the 1000 Genomes Project, an international research initiative that has compiled a catalogue of human genetic variation across the world; the Exome Aggregation Consortium (ExAC) database, a collection of exome sequencing data (the exome is all DNA that codes for proteins); and ClinVar, an NIH database of variants and associated physical conditions. Making sense of the variants is the next stage, and companies have been vigorously bidding for the job. In spring 2014, Genomics England launched a 'bake off' to gauge the quality of genome interpretation expertise from around the world. Twenty-eight participating companies were asked to provide genetic-variant annotation (describing an individual gene and its protein product) and interpretation (assigning them with a function) for the genomes of 15 people with rare diseases compared with 2 samples from healthy relatives, and for 10 genome pairs of tumour DNA and undiseased DNA taken from the same individual.

The ten best-performing companies then took part in a tender to provide annotation and interpretation services for the first 8,000 patients in the project. Genomics England has now narrowed down the pool to four bidders, subject to them passing a test phase and agreeing a contract. Congenica of Cambridge, UK, and Omicia in Oakland, California, will analyse rare-disease genomes; Nanthealth in Culver City, California, will analyse cancer genomes; and WuXi NextCODE in Cambridge, Massachusetts, will analyse both cancer and rare diseases. Cypher Genomics in San Diego, California, and its partner, the advanced technology and defence firm Lockheed Martin in Bethesda, Maryland, are reserve bidders. Most have previous experience in the disease areas for which they were chosen.

All companies will use high-performance computing to interpret genomic data and will work in Genomics England's secure data centre, where the analysis will take place. Their aim is to provide a mostly automated service — interpreting next-generation sequencing data still remains a partly manual process that can take from a couple of hours to weeks. At its peak, says Augusto Rendon, who directs bioinformatics for Genomics England, the project will be receiving up to 200 genomes a day for processing. A manual process will not do if the initiative is to come in on time and on budget — and, ultimately, if whole-genome sequencing is to serve as a widespread diagnosis tool.

THE CHOSEN FEW

Each company brings its own, disparate expertise and history to tackling the thorny interpretation task. Congenica, a spin-off of the Wellcome Trust Sanger Institute and the UK Department of Health, already does testing for NHS laboratories. It will report on inherited and acquired rare-gene mutations associated with disease. The company's Sapientia

THE CLINICAL GENOME

Genomics England plans to sequence 100,000 genomes by 2017. The genomic data will be crucial for diagnosing and treating disease, but its interpretation will require automated, specialized software.

~50,000 people with rare diseases and their parents

~25,000 people with cancer

Tumour DNA

Normal DNA

DNA

RECRUITMENT OF 75,000 PEOPLE
The 100,000 Genomes Project is recruiting people with cancer and rare diseases. The genomes of both normal and tumour cells will be sequenced in people with cancer.

NEXT-GENERATION SEQUENCING
The Californian company Illumina will use UK-based high-throughput sequencing machines to produce whole-genome sequences and identify genetic variants.

AUTOMATED INTERPRETATION
Four UK and US companies will use specialized software to automatically analyse the genetic variants that may be linked to disease.

CLINICAL INTERPRETATION
Around 2,000 UK scientists and clinicians will pore over the data to validate or better understand how the variants may cause disease before the information is fed back to patients.

platform has already been put to the test by analysing DNA from 12,000 children taking part in the Deciphering Developmental Disorders study, the world's largest nationwide rare-disease-sequencing programme[4]. This genome-wide study returned a diagnosis to 30–40% of participating families whose children presented with undiagnosed birth defects or learning disabilities, says Tom Weaver, chief executive of Congenica.

Omicia in the United States also has a track record of providing clinical interpretation, but using open-source, open-access tools. For the 100,000 Genomes Project, it will use gene-interpretation software called Opal to predict which variants are likely to be causing disease. The company's Phevor (phenotype driven variant ontological re-ranking tool)[5] algorithm can also take into account the patient's phenotype — the physical manifestation of disease. These highly automated tools mean that the firm can avoid having to manually compile information from a bundle of sources, says Martin Reese, the company's chief scientific officer. The algorithms statistically 'triage' variants into those that are known to cause disease, those for which disease links exist and those that need further investigation. Opal then pulls together all of the results into a report with which a physician can make treatment decisions.

Cypher Genomics, which is a spin-off of the Scripps Research Institute in La Jolla, California, has developed interpretation software called Mantis that ranks variants according to how likely they are to cause disease. Scientists can then follow up on the tougher cases, which can include manual analysis, says Adam Simpson, the company's chief operating officer.

The only finalist with population-wide genomics know-how under its belt is WuXi NextCODE. The firm is a spin-off of the Icelandic genetics company deCODE, which was founded in 1996 by Kári Stefánsson, who pioneered the idea of linking genealogical and genetic data from the entire Icelandic population to identify human genes associated with common diseases. When deCODE was acquired by the California-based biopharmaceutical firm Amgen, a handful of informaticians and deCODE alumni formed NextCODE, which in turn was acquired by biopharma firm WuXi PharmaTech of Shanghai, China, in January 2015.

WuXi NextCODE has created a way to provide information beyond a distilled list of possible variants. The company has built a way to handle large data sets, such that scientists can zoom in and out of a person's entire genome, and the software continuously pulls out the most up-to-date information about a variant, says Jeff Gulcher, WuXi NextCODE's chief scientific officer.

A typical question might be "I want to know if this is a real mutation or not", says Gulcher. That can mean hunting for other individuals with a rare disease and the same genomic variants. And, one day, a physician might want to find people with cancer who have similar mutations and disease course and who were treated with comparable drugs in the past ten years. The company's Genomically Ordered Relational Database is optimized for such hunts, he says.

In another case, a geneticist might want to compare 20 genomes, each with one million variants, to find out what ails a patient. Rather than compare the genomes in their entirety, WuXi NextCODE's database architecture arranges variants according to genomic position, pulling in a slice of information at a time, making the process computationally efficient. The platform has also been trained on large data sets, such as the genomic information of 300,000 Icelanders, says the company's chief operating officer, Hannes Smarason.

Nanthealth is a privately owned health-care company that is focused on using computational analysis of cancer genomic information to guide diagnosis and therapy. The firm, which did not respond to requests for information, says on its website that it has analysed more than 20,000 genomes. It was founded by physician and biomedical researcher Patrick Soon-Shiong, who also chairs a foundation that funds research and aims to erase disparities in health-care access and heads a non-profit research organization whose goal is to facilitate digital molecular diagnosis. He developed the cancer drug Abraxane (protein-bound paclitaxel), which is used to treat many types of cancers.

For Genomics England, the entire project is about providing clinical answers. But before the diagnostic readouts can reach physicians and their patients, teams of scientists and clinicians will need to pore over the data. Genomics England has organized 2,000 UK scientists with expertise in 13 rare diseases and 10 cancer types to quality check and study project results, especially where a connection to disease has not yet been well established. Using cell-biology assays and mouse models, they will explore how variants may cause or contribute to disease, knowledge that will feed back into the 100,000 Genomes Project. Some results may be quick to validate in databases, whereas others will require careful scrutiny with deeper analysis using the literature, software tools and laboratory assays.

The NHS is not known as a technological innovator, says Bell. But Genomics England will probably be transformative, he says. The 100,000 Genomes Project has generated a wave of activity that could be a powerful boost for the whole field of genomics. It will engender many commercial and academic opportunities — perhaps enabling genomic medicine to finally fulfil its promise of delivering widespread benefit to people with disease. ∎

Vivien Marx *is technology editor for* Nature *and* Nature Methods.

1. Bainbridge, M. N. *et al. Sci. Transl. Med.* **3,** 87re3 (2011).
2. Taylor, J. C. *et al. Nature Genet.* **47,** 717–726 (2015).
3. Raczy, C. *et al. Bioinformatics* **29,** 2041–2043 (2013).
4. Wright, C. F. *et al. Lancet* **385,** 1305–1314 (2015).
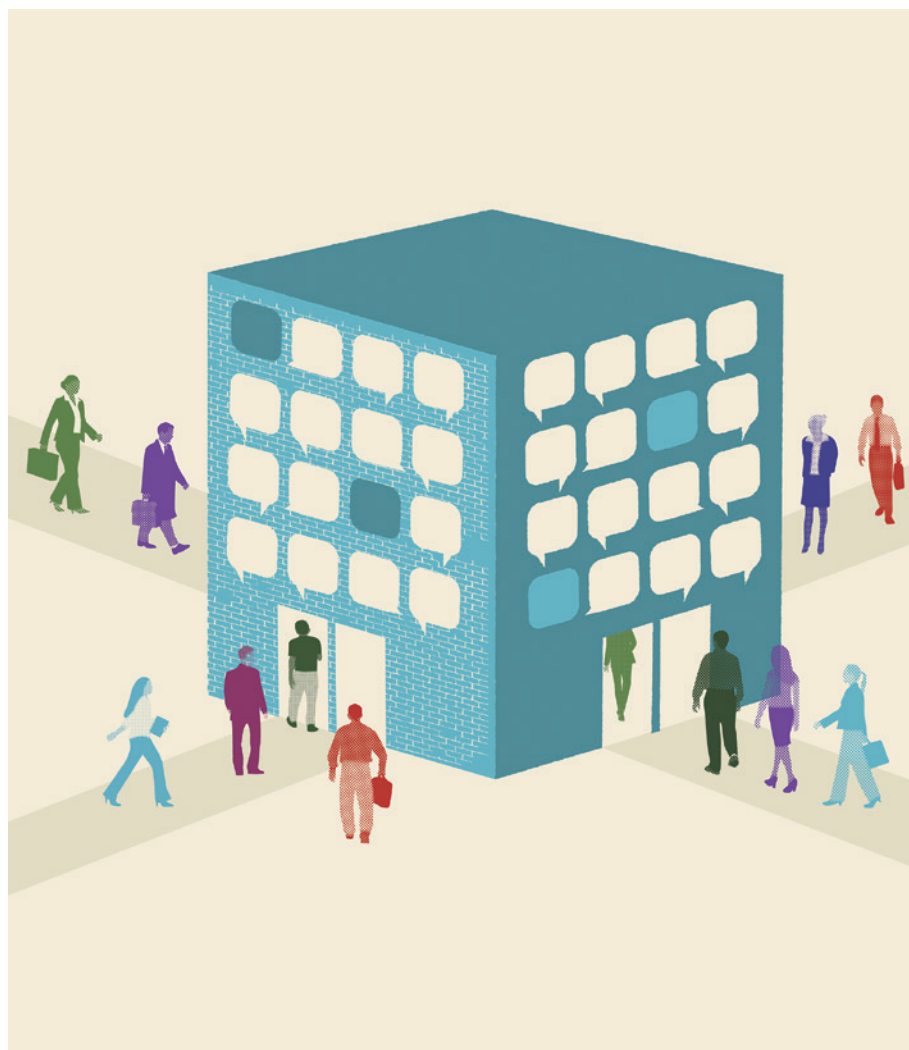5. Singleton, M. V. *et al. Am. J. Hum. Genet.* **94,** 599–610 (2014).

CAREERS

GILLIAN BLEASE/GETTY

COLUMN
Know your network

Seek and cultivate professional relationships to advance your career, says **Peter Fiske**.

Networking is one of the most important activities that a researcher can engage in to develop his or her career. Yet many young scientists are highly anxious about actually doing it. Early-career researchers often fear that networking requires 'schmoozing' strangers, attending tedious meet-and-greet events or relentlessly self-promoting to a degree that would tax even the most outgoing of extroverts. For those who are shy or who do not speak the native language, the prospect of being forced to mix and mingle may seem an insurmountable barrier to career success.

In reality, junior scientists should find professional networking easy and enjoyable. Establishing contacts does not have to involve huge amounts of face time. The key is to aim for a steady level of regular interaction with others — think titration, not chain reaction.

PhD students and postdocs almost universally receive the message from their professors, laboratory heads, mentors and advisers that their publications and research record are the principal means by which they will advance their careers. Those same people tell these researchers that their adviser's colleagues will be the main source of their next professional opportunities. They might implicitly or even overtly discourage students and postdocs from engaging in non-research activities for fear that these extracurricular pursuits could marginally erode the trainee's research productivity. Although this advice is generally well intentioned, it can leave early-career scientists thinking that they have no effective professional network of their own.

SHELL BY SHELL
As a junior researcher, you might not realize that you already have the building blocks for a large and powerful network. A professional network is not simply a list of your colleagues and friends; your network comprises all of the people you know directly and to whom you feel comfortable reaching out with a question or a request. Think of this group as your career 'valence band', or the 'first shell' of your network, populated by the people you know first-hand.

Many young scientists assume that friends and family members in their valence band who are outside the scientific enterprise hold little professional-networking value. In reality, friends and family members have their own networks, and those networks may contain a few people who might be able to help you in your job search or with career development. And because your friends and family members know and care about you, they are often eager to do whatever they can to help you, including warmly introducing you to anyone in their own networks.

The 'second shell' of your network — your friends' friends and contacts — plays a huge part in fostering your career progress and development. For one thing, there are a lot of people at this level. If your immediate network consists of 150 people to whom you feel comfortable asking for help, and each ▶

of them has a similarly sized network, theoretically, you have a 'conduction band' in your network of 22,500 people. At least a few people will be in careers or positions in which they could be of enormous help to you.

It is the importance of this second shell that makes social-networking sites such as Facebook and LinkedIn so valuable: these sites help you to see who is in that shell. You know who your own friends and contacts are, but you have no way of knowing who all your friends' friends and contacts are. Before the launch of sites such as LinkedIn, you would have had to explicitly query all your friends to find out if they knew any-one who worked at a specific organization or company. Now, you can see their net-work links for yourself. Of course, doing so requires that you first set up and cul-tivate your own profile on such sites (see go.nature.com/znl2ea).

Research[1] — including studies that use data from these and other social-networking sites[2] — suggests that although the numbers in the 'third shell' of your network — friends of your friends' friends — are huge, their utility in your career is limited. Third-shell people share no personal connection with you and so are not predisposed to help you. If you want to communicate with a third-shell contact, you should first solidify your rela-tionship with the person in your second shell who connects the two of you. In effect, you are turning the second-shell contact into a first-shell friend. How do you do that? Build a relationship with that person through cor-respondence or by meeting them in person, if possible. An informational interview could work in this case.

NURTURE YOUR NETWORK

The development of a professional network does not in and of itself immediately solve your networking needs. You must effec-tively manage and cultivate your network, no matter how large or small it may be.

Junior researchers need to remember that their professional network is nothing more than a set of personal relationships.

And personal relationships thrive with a bit of attention. Some people send out peri-odic updates to members of their network (we used to call these letters). Some post updates on social-networking sites. And others prefer to maintain and nurture their network by calling one person each day to whom they have not spoken in a while. This 'depth versus breadth' strategy might help you to learn about important news, updates and leads.

Of course, a network is not simply a set of relationships to be tended, but also a specific and powerful tool to use both during a job search and throughout your career. When you learn of an ideal job opportunity, you can ask members of your first-shell network for a personal introduction to someone whom they know in that organization. Or, if you are seeking an oppor-tunity at an organi-zation, institute or company that is not advertising an open position, you can reach out directly to someone in your second shell who works there and ask that person about his or her employer. Better yet, set up an infor-mational interview with that person. Face time almost always yields better results than does electronic communication or even a phone call.

> "The degree to which you help others is often linked to how much help you yourself receive."

You may worry about asking for too many favours from members of your con-tacts' network. But if you feel embarrassed to ask for help, or fear that someone will be offended or disappointed if you ask, get over it. Often, a member of your network will be more disappointed if they learn that they missed an opportunity to help you because you failed to reach out.

And remember that your network operates in two directions: the degree to which you help others is often linked to how much help you yourself receive. Your network becomes stronger through the help that you give.

A well-tended and extensive network is one of the most valuable assets for pro-fessionals in today's economy. Those who invest in both their work and their rela-tionships will reap the greatest number of opportunities. ∎

Peter Fiske *is chief executive of PAX Water Technologies in Richmond, California, and author of* Put Your Science to Work *(American Geophysical Union, 2001).*

1. Travers, J. & Milgram, S. *Sociometry* **32**, 425–443 (1969).
2. Kietzmann, J. H., Hermkens, K., McCarthy, I. P. & Silvestre, B. S. *Bus. Horizons* **54**, 241–251 (2011).

TRADE TALK
Impact assessor

Arie Meir works at Google.org, the philanthropic arm of the technology giant in Mountain View, California. He and his team evaluate grant proposals for technical feasibility and their level of social impact. Here, he describes his path from a PhD in biophysics to team leadership at Google.

Why did you go to graduate school?
I was programming for GE Healthcare, and I realized that everything I was doing was the result of someone else's intellectual process — the product of someone else's thinking. I wanted to get to the source. I thought that if I could learn to do scientific research, that toolkit would enable me to apply those skills to a lot of other problems.

How did your training help you to get this job?
I gained exposure to other careers. I probably did 100 informational interviews. When you say you're a student from a university that people know, it opens doors. For example, I didn't know there was a field called manage-ment consulting until I met someone who was doing it.

Which skills have been the most useful?
I learned how to solve complicated problems and to communicate in a structured way. I know how to convert an abstract idea into practice and how to design an experiment. It doesn't matter what job you do; you will use this skill set. Another thing was strategic-skill development: how to position myself for opportunities. When I learned about technolo-gies such as 3D printing, I also became curious about social entrepreneurship. I was talking with friends about it, and one from Colombia told me, 'I have a non-profit; go talk to them'. So, while I was working on my PhD, I launched a small non-profit organization focused on technical education in Latin America.

Is that how you came to join Google.org?
That was just an experiment, but I used the experience later, during my job interview with Google.org. I knew how to speak the language that the people on my team speak. Then I became relevant — it was not, 'Arie is a scientist'. It was, 'Arie is one of us'. ∎

INTERVIEW BY MONYA BAKER

This interview has been edited for length and clarity; see go.nature.com/y7qxcg for more.

UNDER AN UNCARING SKY

Ill met by starlight.

BY WILLIAM MEIKLE

I volunteered for this — our ore would last only ten more years, so somebody had to do something. The flyer they gave me hadn't been upside for 30 years; nobody had. The dimming had seen to that. Too cold, too dark. No point.

I switched on the ore probe and let it run. I had nothing to do for hours except hang there in the sky and try to ignore the stars that now seemed to be falling ever closer, threatening to wrap themselves around me, engulf me and drag me off to the black beyond. I almost didn't hear the *beep* as the probe finally announced a finding.

The flyer told me I was somewhere in the South Pacific. It looked little different from the spot where I had come upside, but as I descended I saw that the ice here was less compacted. Several darker patches showed. As I got closer, I could see that there were stretches of broken ice and slush. I started to think there might even be open water available. The probe beeped a proximity alert as the flyer hovered ten feet above an island of rock, black against the ice all around.

There was ore here, and a lot of it. The scan showed a seam, some 100 feet deep in the rock. I put the flyer down on the flattest spot I could find and started the drill before I put on a helmet and ventured outside, aware even as I did so that I was probably the first human to walk above ground for three centuries.

The only sound was the steady grinding of drill on rock. My heads-up told me that the strata being drilled were sedimentary on top of schist, the drill currently penetrating rock that was more than two hundred million years old, and going through a million years of sediment a second.

All of which was secondary to the fact that I had just found a cave.

I almost felt at home. The walls were smooth — some weathering process over the millennia was my assumption — and the light from my helmet was bright enough to light my way for 20 yards ahead. The cave floor sloped downwards, and as I proceeded the temperature rose. Fifty yards in, I had to turn off the suit heater. I also got the first indication that this was more than just a simple cave. I found a number imprinted on the wall:

SUB-LEVEL 25.

The passageway was man-made.

I found the first corpse seconds later.

I did not look too closely, but judging by the decomposition of the clothing, I guessed that many years had passed since the time of death. I kept going, but I was no longer hopeful that I would meet anyone still alive.

The corridor opened into a wider chamber, an eating area of sorts.

Bodies lay strewn everywhere, lying in mounds or in pairs. Skeletal arms were wrapped around broken necks, skulls showed signs of having been smashed against table and floor. They had killed each other in a frenzied melee.

As I bent to inspect the closest, I saw the cause.

The darkness danced in their eye sockets, a deeper shadow. It was full of stars where the sky had fallen in and got them. The more I looked, the more I saw it; there in the shadow where a body hung over an overturned chair, there in the corner under the food processors, but

ILLUSTRATION BY JACEY

mostly in the eyes, dancing and twinkling, mocking my horror.

I stumbled past more bodies than I could count, searching for a reason, an answer as to what had happened. The empty eyes followed me everywhere I went. There was a door opposite me, and I went through, hoping for some small escape from the terror.

There were no bodies in this part of the system, just long empty corridors, but that somehow only made matters worse. Although I was still safe inside the suit, the air seemed somehow thicker here, more oppressive; a faint trace of blue mist hung in the air. If I were back home, I knew I would be approaching the bionic plant now. Despite the terrors of the eating area above, I was almost eager to visit the working parts of the site, as there might even be something salvageable there, something that would further prolong our own time here in the warren.

I descended a stairwell and walked out into their bionics plant.

Scores of eyes turned and looked at me, reflecting like twinkling stars in my helmet light.

They had once been human, but their skin was pale, almost translucent, their eyes large, like saucers in heads too small to hold them. They scrambled, on all fours, amid a pile of slurry and started to crawl towards me, piteously mewling like hungry kittens. Stars danced in their eyes.

I will not tell of my flight from that place, save to say that I have deleted the coordinates from the systems. If you want the ore, you will have to send out another flyer.

But I would advise against it, for the darkness will come back with them. The sky will fall, and your eyes will fill with stars. The darkness will get inside, and it will consume you, as it did those poor things in the bionics lab … as it has started to do to me.

It is vast, it is empty, and it does not care.

It just does not care. ∎

William Meikle *is a Scottish writer, now living in Canada, with 20 novels published in the genre press and more than 300 short-story credits in 13 countries.*

Questioning evidence of group selection in spiders

Any field study showing convincing evidence of group selection would be a significant contribution to the field of evolutionary biology. Pruitt and Goodnight[1] claim to provide such evidence in a 14–18-month field experiment on spiders. However, we contend that apparent flaws in their predictions, assumptions, methods and interpretations undermine this claim. We believe that the data presented are unreliable and are equally consistent with both group selection and individual-level selection; thus, we question the conclusion of Pruitt and Goodnight[1] that group selection has produced the observed patterns. There is a Reply to this Brief Communication Arising by Pruitt, J. N. & Goodnight, C. J. *Nature* **524**, http://dx.doi.org/10.1038/nature14597 (2015).

Evaluating group selection involves, at a minimum, estimating and comparing both individual and group fitness, as stated by previous reviews[2,3] and performed by other studies[4,5]. However, Pruitt and Goodnight[1] did not estimate individual fitness, and so cannot evaluate the relative importance of group selection compared to individual-level selection. The chosen species, *Anelosimus studiosus*, is solitary, rarely forms groups[6], and shows no evidence of reproductive restraint

or skew within groups[7]. Thus, individual and group fitness are not expected to conflict and are generally confounded, emphasizing how crucial it is, first, to formulate predictions capable of distinguishing individual-level selection and group selection explanations and, second, to estimate individual fitness.

Both predictions of Pruitt and Goodnight[1] could follow equally well from individual-level selection as from group selection. Their first prediction is "compositions that approximate the normal mixtures that characterize each site will enjoy greater success", where 'compositions' refers to within-group phenotypic frequencies. Merely demonstrating differential survival of groups does not allow the authors to distinguish successful groups from groups of successful individuals. The 'group trait' of Pruitt and Goodnight[1] is a group-size-dependent behavioural polymorphism. Experimental changes in this group trait (that is, manipulating group size and phenotype frequency) may directly affect within-group individual fitness just as well as whole-group fitness[8–11]. Specifically, creating experimental groups that deviate from locally stable polymorphisms may reduce mean individual fitness, rendering group extinction more likely. The prediction of
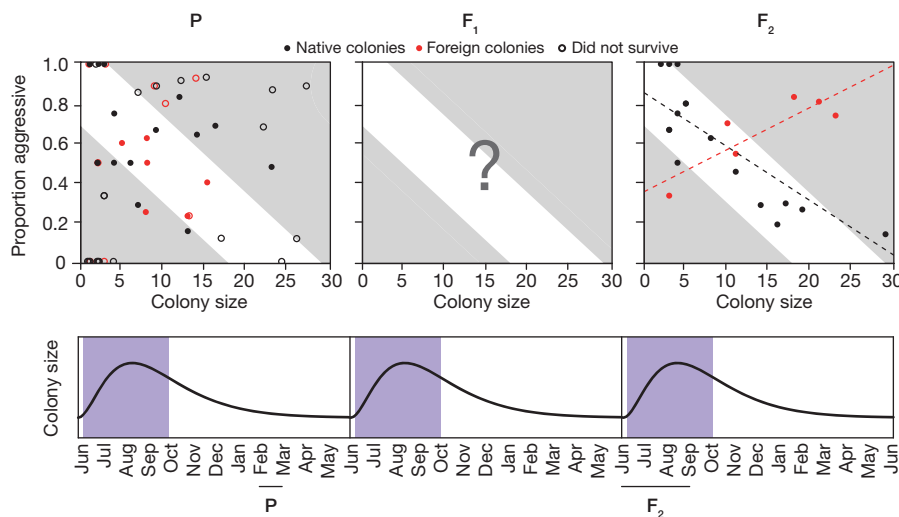

Figure 1 | Overview of methods and results from Pruitt and Goodnight[1]. The top panel shows the distribution of experimental colonies placed in all six field sites in the parental (P) generation; the missing information of the next generation (F₁); and the distribution of final compositions of the grandoffspring generation (F₂). Phenotypic compositions, that is, proportions of aggressive individuals in each colony, are plotted against colony sizes. We present data only from one low-resource field site, Don Carter, to illustrate the setup. Black dots represent native colonies (created with spiders collected at Don Carter); red dots represent foreign colonies (spiders collected at high-resource field site Moccasin Creek). Filled circles of both colours in P are colonies that were still alive in F₂ (equal to figure 1c of Pruitt and Goodnight[1]); empty circles are colonies that had gone extinct by F₂. The white band represents the proposed selection pressure at that field site: a regression line fitted on phenotypic compositions and colony sizes of naturally occurring colonies at Don Carter (based here on colonies of sizes up to 30; its thickness chosen arbitrarily). Dotted lines in F₂ represent regressions of the final F₂ compositions of the surviving colonies: native (black) versus foreign (red). Surviving colonies had P compositions close to the white selection band, but F₂

compositions differed according to site of origin: Native F₂ compositions were close to the selection band while foreign F₂ compositions followed a positive regression, dissimilar to the selection band. Setup and results were similar in the two additional low-resource sites while the three high-resource sites showed opposite trends (that is, selection bands were positive regression lines while the foreign F₂ regressions showed negative correlations). Note that although foreign colonies end up opposite to the proposed selection pressure, Pruitt and Goodnight[1] still conclude that their proposed selection pressure was supported. The bottom panel shows how the size of an *A. studiosus* colony is expected to vary within years with a peak around egg hatching during summer. The period of maternal care is marked in purple. Black lines marked with 'P' and 'F₂' indicate at which point in the life cycle Pruitt and Goodnight[1] performed the behavioural assays to determine the phenotypic compositions of colonies. Note that group sizes and phenotypic compositions of P and F₂ were measured at different points, apparently comparing sexually mature females (P) with juvenile grandchildren (F₂) during maternal care at a stage where offspring sex cannot be determined.

differential group extinction can therefore result from individual-level selection just as plausibly as from group selection. Similarly ambiguous is the second prediction that "colonies should only be able to adaptively hone compositions when composed of native individuals". If 'native colonies' can 'adaptively' change phenotype frequencies over time, this may occur via several mechanisms, as Pruitt and Goodnight[1] mention (plasticity, phenotype-biased dispersal, and so on). Yet, any of these mechanisms may evolve by individual-level selection, a possibility overlooked by Pruitt and Goodnight[1].

The conclusions of Pruitt and Goodnight[1] rest on the assumption that 'naturally occurring mixtures' (that is, field phenotypic frequencies; see figure 1a of Pruitt and Goodnight[1]) represent consistent selection pressures across years. Yet, the years of measurement were patchy (2007–14), differed among sites and often did not overlap (Table 1). Indeed, Pruitt and Goodnight[1] sampled significantly different phenotypic mixtures and group sizes among years at each site (mixtures: $P = 1.1 \times 10^{-5}$; group size: $P = 7.8 \times 10^{-14}$; Fisher's combined P value across separate Kruskal–Wallis tests for each site) but overlooked this variation and pooled dissimilar data. In four out of nine samples taken at high resource sites, the selection pressure was no different from zero (non-significant correlations between mixtures and group size; separate *lm* for each year at each site), and in two out of the three low resource sites the relationship between mixture and group size differed significantly from year to year (Norris Dam: $P = 0.0074$; Don Carter: $P = 0.017$; interaction between log(group size) and year on phenotypic mixtures in linear models). Moreover, half of the sites had not been assessed for four to six years. These measurements cannot be assumed to represent consistent, current selection pressures.

Pruitt and Goodnight[1] infer group selection by comparing parental (P) and grand-offspring (F$_2$) generations based on parental traits alone: they compare P-phenotypic compositions of P colonies with P (not F$_2$) compositions of surviving F$_2$ colonies (figure 1c of Pruitt and Goodnight[1] is identical to their figure 1b, minus extinct colonies: F$_2$ colonies are depicted with their grandparents' compositions). We question the validity of assuming that past compositions are visible to selection but present compositions are not. Indeed, 'foreign colonies' changed to display F$_2$ compositions in a pattern opposite to the assumed selection pressure (Fig. 1; see figure 2 of Pruitt and Goodnight[1]). These changes mean that F$_1$ compositions presumably also differed from P compositions (and were visible to selection during that generation) but F$_1$ was not assessed (Fig. 1).

Changes within generations were also not considered (Fig. 1). Around egg hatching[12], colonies peak in size, after which mortality and dispersal decrease colony size. Phenotypic composition cannot be considered a stable 'group trait' when its proposed selective advantage is a function of group size, and group size changes nonlinearly over time. Compounding this, compositions of P and F$_2$ were apparently measured at different developmental stages (Fig. 1): we believe that this is a serious flaw, as individual phenotypes are affected by reproductive status[13]. Pruitt and Goodnight[1] compared sexually mature females (P) with grandchildren (F$_2$) that, given the stated timescale, presumably were juvenile, mixed-gender and receiving maternal care (Fig. 1; mothers die off in October[12] and juvenile spiders are unsexable).

We believe that none of the findings of Pruitt and Goodnight[1] supports their claim to have demonstrated a "marked evolutionary response to group selection". Rather, after two generations, surviving foreign colonies failed to change phenotypic compositions in site-appropriate ways (instead changing to express compositions appropriate for their original site), suggesting a lack of genetic change over the experiment (Fig. 1). Pruitt and Goodnight[1] argue that this constitutes evidence that mechanisms for adjusting compositions are locally adapted due to historical group selection, but provide no justification for this claim: while they provide data suggesting phenotypes themselves may be partially heritable, there is no evidence that this 'adjustment mechanism' has undergone genetic change and, again, no attempt to reject individual-level selection as an explanation. Foreign colonies may revert to their native phenotype compositions without genetic change; for example, due to persistent maternal or epigenetic effects, either of which may respond to individual-level selection or group selection. The evolutionary mechanisms shaping population-level differences, whether in phenotype frequencies or the means by which these change over time, do not appear to have been addressed in this paper.

We would welcome any field study demonstrating that group selection causes genetic change over generations in ways inconsistent with individual-level selection. Given recent high-profile exchanges over the relative importance of group selection[14,15], such a paper would be a significant contribution to the field. Unfortunately, we do not believe that the paper of Pruitt and Goodnight[1] is such a study.

Lena Grinsted[1], Trine Bilde[2] & James D. J. Gilbert[1,3]
[1]School of Life Sciences, University of Sussex, John Maynard Smith Building, Brighton BN1 9QG, UK.
email: l.grinsted@sussex.ac.uk
[2]Department of Bioscience, Aarhus University, Ny Munkegade 116, building 1540, 8000 Aarhus C, Denmark.
[3]School of Biological, Biomedical and Environmental Sciences, University of Hull, Cottingham Road, Hull HU6 7RX, UK.

1. Pruitt, J. N. & Goodnight, C. J. Site-specific group selection drives locally adapted group compositions. *Nature* **514,** 359–362 (2014).
2. Eldakar, O. T. & Wilson, D. S. Eight criticisms not to make about group selection. *Evolution* **65,** 1523–1526 (2011).
3. Marshall, J. A. R. Group selection and kin selection: formally equivalent approaches. *Trends Ecol. Evol.* **26,** 325–332 (2011).
4. Rainey, P. B. & Rainey, K. Evolution of cooperation and conflict in experimental bacterial populations. *Nature* **425,** 72–74 (2003).
5. Kerr, B., Neuhauser, C., Bohannan, B. J. M. & Dean, A. M. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature* **442,** 75–78 (2006).
6. Grinsted, L. & Bacon, J. P. Animal behaviour: task differentiation by personality in spider groups. *Curr. Biol.* **24,** R749–R751 (2014).
7. Jones, T. C., Pruitt, J. N. & Riechert, S. E. Reproductive success in a socially polymorphic spider: social individuals experience depressed reproductive success in isolation. *Ecol. Entomol.* **35,** 684–690 (2010).
8. Moczek, A. P., Hunt, J., Emlen, D. J. & Simmons, L. W. Threshold evolution in exotic populations of a polyphenic beetle. *Evol. Ecol. Res.* **4,** 587–601 (2002).
9. Culumber, Z. W., Bautista-Hernandez, C. E., Monks, S., Arias-Rodriguez, L. & Tobler, M. Variation in melanism and female preference in proximate but ecologically distinct environments. *Ethology* **120,** 1090–1100 (2014).
10. Weider, L. J. *et al.* Genotype × environment interactions, stoichiometric food quality effects, and clonal coexistence in *Daphnia pulex*. *Oecologia* **143,** 537–547 (2005).

Table 1 | Collection years for each field site

Site	Collection years							
	2007	2008	2009	2010	2011	2012	2013	2014
High resource sites								
Melton Hill	X	X	X	E	E X	X		
Little River	X	X					E	E
Moccasin Creek							E X	E X
Low resource sites								
Norris Dam	X	X	X	E	E X	X		
Clinch River		X		X			E	E
Don Carter			X	X			E	E

Data from years marked with 'X' were pooled within sites to create the naturally occurring mixtures (figure 1a in Pruitt and Goodnight[1]) used to infer site-specific selection pressures. 'E' indicates the years during which the experiment took place.

BRIEF COMMUNICATIONS ARISING

11. Fitzpatrick, M. J., Feder, E., Rowe, L. & Sokolowski, M. B. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature* **447**, 210–212 (2007).
12. Furey, F. E. Two cooperatively social populations of the theridiid spider Anelosimus studiosus in a temperate region. *Anim. Behav.* **55**, 727–735 (1998).
13. Watts, J. C., Ross, C. R. & Jones, T. C. Diel and life-history characteristics of personality: consistency versus flexibility in relation to ecological change. *Anim. Behav.* **101**, 43–49 (2015).
14. Abbot, P. *et al.* Inclusive fitness theory and eusociality. *Nature* **471**, E1–E4 (2011).
15. West, S. A., Griffin, A. S. & Gardner, A. Social semantics: how useful has group selection been? *J. Evol. Biol.* **21**, 374–385 (2007).

Group selection versus group adaptation

Pruitt and Goodnight[1] describe how the ratio of aggressive versus docile females varies among naturally occurring colonies of the social spider *Anelosimus studiosus*, with larger colonies exhibiting more aggression in high-resource environments and the reverse in low-resource environments. They experimentally manipulate this ratio to show that it influences a colony's reproductive success. Pruitt and Goodnight[1] conclude that this work demonstrates group-level adaptation and contradicts an earlier theoretical analysis[2]. Here, I show that this conclusion is unfounded and arises from a conceptual misunderstanding. There is a Reply to this Brief Communication Arising by Pruitt, J. N. & Goodnight, C. J. *Nature* **524,** http://dx.doi.org/10.1038/nature14597 (2015).

While Pruitt and Goodnight[1] provide evidence of group-level selection, they do not provide any evidence of group-level adaptation, as defined in the earlier analysis[2]. A response to group-level selection occurs when there is heritable variation in group fitness and—along with selection acting within groups—this may contribute to evolutionary change[3,4]. Owing to the mathematical equivalence of multilevel-selection and kin-selection analysis, this is entirely consistent with individuals being adapted to maximize their inclusive fitness[3]. In contrast, group-level adaptation is the stronger notion that phenotypes are optimized for the good of the group, a design objective that is typically in conflict with the individual's inclusive-fitness interests and which will rarely be favoured by natural selection[2,5]. Other definitions of group-level adaptation are possible, but this is the definition given in the earlier analysis[2] that Pruitt and Goodnight[1] claim to have refuted. Pruitt and Goodnight[1] present evidence that levels of aggression have been, at least in part, moulded by group-level selection, but they do not show that this ratio is optimized for the good of the group.

To illustrate this point, I adapt Frank's[6] 'tragedy of the commons' model to study a scenario where aggressive individuals are competitively superior within groups but an intermediate level of aggression, depending on colony size and resource availability, is favoured at the group level (see Methods). The resulting level of aggression favoured by natural selection: (1) is that which balances within-group and between-group selection pressures and, accordingly, maximizes the individual's inclusive fitness; (2) increases with colony size in high-quality environments and decreases with colony size in low-quality environments, in line with the data of Pruitt and Goodnight[1]; and (3) is generally higher than that which maximizes group fitness (Fig. 1). Accordingly, the data presented by Pruitt and Goodnight[1] neither invalidate the idea that individuals are adapted to maximize their inclusive fitness nor validate the idea that the colony is adapted to maximize its own fitness.

Pruitt and Goodnight[1] also claim that the earlier theoretical analysis[2] had suggested that group-level adaptation can occur only in the context of clonal groups. This is incorrect, as the analysis made clear that group-level adaptation can occur in genetically heterogeneous groups, as long as there is a mechanism for suppressing within-group conflict[2]. Well-studied examples of such mechanisms are fair meiosis[7,8] and worker policing in honeybees[8,9]. Interestingly, Pruitt and Goodnight[1] suggest that policing may occur in *A. studiosus* colonies. If true, then there is scope for group-level adaptation, but further study would be needed to confirm this.

These points echo remarks made by Maynard Smith[5], in connection with a different phenotypic polymorphism: females versus males. The sex ratio is perhaps the best-studied social-evolutionary trait, and provides some of the best quantitative evidence for Darwinian adaptation in the natural world[10]. Certain female-biased sex ratios are recognized to be driven, in part, by selection acting at the level of the group[11,12], and experimental manipulations have confirmed the impact of group sex ratio on group fitness[13]. But Maynard Smith[5] cautioned that, because the sex ratio that evolves is not that which maximizes group fitness, but rather that which balances within-group and between-group selection pressures, it does not constitute a group-level adaptation. Instead, it represents the adaptation of individual organisms, for the purpose of maximizing their inclusive fitness.
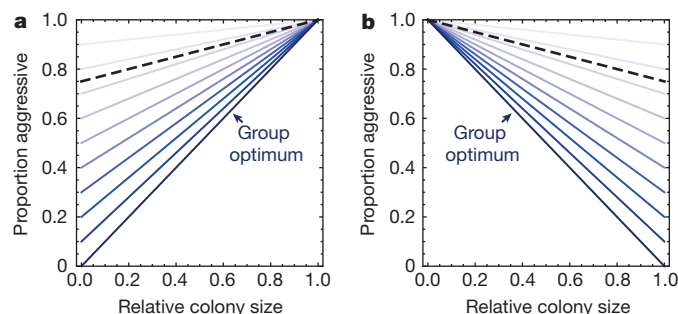

Figure 1 | Individual versus group optima. The proportion of aggressive individuals resulting from the maximization of the individual's inclusive fitness is plotted for a range of coefficients of relatedness, from $r = 0$ (lightest shading) to $r = 1$ (darkest shading), with the estimate of Pruitt and Goodnight[1] $r = 0.25$ marked as a dashed line. The group optimum corresponds to the $r = 1$ line; that is, when natural selection is acting only at the level of the colony. **a**, In high-resource environments, the level of aggression that maximizes the individual's inclusive fitness increases with colony size, in line with the data of Pruitt and Goodnight[1], and typically exceeds that which maximizes colony fitness. **b**, In low-resource environments, the level of aggression that maximizes the individual's inclusive fitness decreases with colony size, in line with the data of Pruitt and Goodnight[1], and typically exceeds that which maximizes colony fitness.

Methods

A female's fitness is $w = fg$, where $f = x/y$ describes her within-group advantage if she is aggressive with probability x and the average female in her colony is aggressive with probability y; and $g = y^b(1 - y)^{1-b}$ describes her colony's fitness.

BRIEF COMMUNICATIONS ARISING

In high-resource environments $b = n$, where n denotes relative colony size, reflecting the finding of Pruitt and Goodnight[1] that aggression is relatively more important for colony survival in this setting; and in low-resource environments $b = 1 - n$, reflecting the finding of Pruitt and Goodnight[1] that the opposite is true in this setting. In both cases, the intermediate inclusive-fitness optimum $z*$ satisfies $((\partial w/\partial x) + r(\partial w/\partial y))|_{x = y = z*} = 0$ (ref. 6).

Andy Gardner[1]

[1]School of Biology, University of St Andrews, Dyers Brae, St Andrews
KY16 9TH, UK.
email: andy.gardner@st-andrews.ac.uk

1. Pruitt, J. N. & Goodnight, C. J. Site-specific group selection drives locally adapted group compositions. *Nature* **514**, 359–362 (2014).
2. Gardner, A. & Grafen, A. Capturing the superorganism: a formal theory of group adaptation. *J. Evol. Biol.* **22**, 659–671 (2009).
3. Hamilton, W. D. In *Biosocial Anthropology* (ed. Fox, R.) 133–155 (Wiley, 1975).
4. Okasha, S. *Evolution and the Levels of Selection* (Oxford Univ. Press, 2006).
5. Maynard Smith, J. In *The Latest on the Best: Essays on Evolution and Optimality* (ed. Dupré, J.) 147–149 (MIT, 1987).
6. Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, 1998).
7. Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proc. Natl Acad. Sci. USA* **74**, 4542–4546 (1977).
8. Frank, S. A. Repression of competition and the evolution of cooperation. *Evolution* **57**, 693–705 (2003).
9. Wenseleers, T. & Ratnieks, F. L. W. Enforced altruism in insect societies. *Nature* **444**, 50 (2006).
10. West, S. A. *Sex Allocation* (Princeton Univ. Press, 2009).
11. Hamilton, W. D. Extraordinary sex ratios. *Science* **156**, 477–488 (1967).
12. Frank, S. A. Hierarchical selection theory and sex ratios. I. General solutions for structured populations. *Theor. Popul. Biol.* **29**, 312–342 (1986).
13. Reece, S. E., Drew, D. R. & Gardner, A. Sex ratio adjustment and kin discrimination in malaria parasites. *Nature* **453**, 609–614 (2008).

Pruitt & Goodnight reply

In Pruitt and Goodnight[1] we provided experimental evidence that group selection has contributed to a group-level adaptation in the social spider *Anelosimus studiosus*. Grinsted *et al.*[2] provide a wide diversity of system-specific critiques of our original study. In contrast, Gardner[3] highlights differences between our definition of 'group level adaptations' and his own. He further describes a model that recreates some of the dynamics seen in *Anelosimus studiosus*. Below, we address the critiques of Grinsted *et al.*[2] first and Gardner[3] second.

Grinsted *et al.*[2] claim that documenting group selection requires that one rule out evidence of individual selection; we believe that this is incorrect. Group selection occurs when there are differences among groups in their survival or reproductive output as a consequence of their traits[4–6], which is what we showed in our paper[1]. It is widely agreed that individual selection, group selection and kin selection act simultaneously in most societies. The controversy is whether group selection contributes significantly to adaptations in the wild[7,8]. We provided evidence that group selection has contributed to a group-level adaptation in *Anelosimus studiosus*[1]. Our case study is clear because both the target and agent of selection are above the level of the individual: the target of selection (group composition) is a trait that an individual cannot have, and the agent of selection (extinction) is the textbook example of strong group selection[9–11]. We showed that *A. studiosus* colonies live or die as a unit[11] because of their behavioural composition[1]. None of the criticisms of Grinsted *et al.*[2] weakens this claim.

Grinsted *et al.*[2] argue that individual traits could underlie colony extinction events—we agree with this assertion. All collective traits can be decomposed into the traits of constituents. A group cannot perish unless the individuals within it die too. Grinsted *et al.*[2] merely want us to focus on a different, non-mutually exclusive, level of analysis. Following their logic, all behavioural studies would be flawed because behaviour can be decomposed into physiology, genetics, applied physics, and so on. Thus, the arguments of Grinsted *et al.*[2] aren't against group selection per se, but instead it seems they take issue with the word 'group'.

Grinsted *et al.*[2] argue that the interests of individuals and groups are united in *A. studiosus*, and we agree. This is because group selection (extinction) is the major force driving individual fitness. Grinsted *et al.*[2] make reference to group selection requiring reproductive skew, and this would appear to suggest that they are conflating group selection with altruism[7]—these are different concepts. There is controversy whether group selection can beget altruism in the absence of relatedness[12], which our data don't address. Yet, there is also interest in understanding the intensity of group selection acting in nature and how group selection varies across environments[4,13], which is the subject of our paper.

Grinsted *et al.*[2] demonstrate that the naturally occurring relationship between group composition and group size at each site differs across years. In particular, they note that the significance of the relationship between group size and composition vanishes in four out of our eighteen observations. Yet, with only one exception, the estimated relationships are always positive for high resource sites and negative for low resource sites. These trends are robust in spite of the inherent variability encountered by any field study. Grinsted *et al.*[2] fail to address the key issue of how our supposedly 'weak' baselines accurately predict colony survival. If these empirical relationships were truly unreliable then they would not accurately predict anything, thus the criticism of Grinsted *et al.*[2] is unfounded.

Grinsted *et al.*[2] claim that one must watch selection at each generation to document selection on group composition. However, that would disrupt the very processes that we were quantifying. Our results demonstrate unequivocally that the composition of the parental generation (P) predicts the number of grand-offspring (F_2) produced by colonies at all six sites. One rarely obtains a higher calibre field estimate of fitness than that.

Grinsted *et al.*[2] claim that we measured spiders at different developmental stages. In fact, we only measured mature females throughout our study. While our colonies had an accelerated phenology, we showed that this does not impact their performance[1]. Aggressiveness varies with temperature[14] and gravidity[15] in *A. studiosus*, which could be problematic. Thankfully, rank order aggressiveness is maintained in spite of this plasticity[14], thus allowing reliable assignment of either phenotype. The criticism of Grinsted *et al.*[2] that *A. studiosus* typically live solitarily is true but irrelevant. *Anelosimus studiosus* typically live in multi-female colonies at these sites[16]. Regardless, this wouldn't compromise any of our findings.

BRIEF COMMUNICATIONS ARISING

Grinsted *et al.*[2] assert that we did not provide any evidence of genetic change as a consequence of group selection. We argue that we provided evidence that the mechanisms used by colonies to adjust their compositions are site specific, genetically influenced, and locally adapted because of group selection. First, we showed that the behaviour of colonies is adaptive: colonies always hone their compositions in ways that would evade extinction at their home sites, which cannot be purely coincidental. Second, we showed that these regulatory mechanisms are genetically influenced: after two generations displaced colonies still exhibited the same regulatory behaviour that they would at their home site. This conveys that the differences in the regulatory behaviour of native/foreign colonies are unlikely to be plastic. Such was the case at every site: six two-generation common garden experiments. Common garden experiments are the gold standard for verifying a genetic component to any phenotype[17].

Grinsted *et al.*[2] claim that they await a study that demonstrates that group selection has caused genetic change in a way that is inconsistent with individual selection; however, those papers have already been published[4,18]. The novelty of our study is that we showed that the nature of group selection changes across environments, and we provide evidence that historic group selection has caused local adaption in a collective trait.

We define 'group-level adaptions' as adaptations in group-level traits that evolve via a combination of individual selection and group selection. Group composition in *A. studiosus* meets these criteria[1,3]. Gardner[3] adds the requirement that 'group-level adaptations' must evolve because of the advantages they confer to the group, at the cost of the individual[3,19]. This definition is problematic not only because it confounds adaptions in group-level traits with the concept of altruism, but also mandates that a trait can only be considered a group-level adaptation if it is optimized solely for the group. However, no trait in the real world is ever the result of adaptation to a single selection pressure. We expect no different from group traits. Group traits are inevitably a compromise among competing selection pressures at the individual and group level. We argue that if one demands that traits can only be counted as adaptations if they are subject to a single selective pressure, then it is almost certainly true that there are no such things as adaptations.

Gardner's model[3] recreates a subset of the patterns seen in *A. studiosus*, although it invokes group selection to do so. While there is nothing wrong with Gardner's interesting model, it's not grounded in the biology of our system. For instance, aggressive individuals are not competitively superior to docile individuals in *A. studiosus*[20], but Gardner's model relies on this assumption[3]. Gardner's model also fails to predict key findings. For example, his model does not explain why displaced colonies recreate the ideal mixtures of their home sites after generations at a new site. In short, an accomplished theorist like Gardner could devise many models that capture some of the patterns seen in our system. We devised one such model ourselves. The difference is that we then vetted our model with arguably the most comprehensive field assessment of group selection ever conducted[1]. Our findings yielded strong support for our model, providing a high standard of scientific evidence.

Jonathan N. Pruitt[1] **& Charles J. Goodnight**[1]
[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15216, USA.
email: pruittj@pitt.edu

1. Pruitt, J. N. & Goodnight, C. J. Site-specific group selection drives locally adapted colony compositions. *Nature* **514,** 359–362 (2014).
2. Grinsted, L., Bilde, T. & Gilbert, J. D. J. Questioning evidence of group selection in spiders. *Nature* **524,** http://dx.doi.org/10.1038/nature14595 (2015).
3. Gardner, A. Group selection versus group adaptation. *Nature* **524,** http://dx.doi.org/10.1038/nature14596 (2015).
4. Goodnight, C. J. & Stevens, L. Experimental studies of group selection: What do they tell us about group selection in nature? *Am. Nat.* **150,** S59–S79 (1997).
5. Goodnight, C. On multilevel selection and kin selection: Contextual analysis meets direct fitness. *Evolution* **67,** 1539–1548 (2013).
6. Eldakar, O. T. & Wilson, D. S. Eight criticisms not to make about group selection. *Evolution* **65,** 1523–1526 (2011).
7. West, S. A., Griffin, A. S. & Gardner, A. Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20,** 415–432 (2007).
8. West, S. A., Griffin, A. S. & Gardner, A. Social semantics: How useful has group selection been? *J. Evol. Biol.* **21,** 374–385 (2008).
9. Wade, M. J. Critical-review of models of group selection. *Q. Rev. Biol.* **53,** 101–114 (1978).
10. Maynard Smith, J. Group selection. *Q. Rev. Biol.* **51,** 277–283 (1976).
11. Lewontin, R. C. The units of selection. *Annu. Rev. Ecol. Syst.* **1,** 1–18 (1970).
12. Abbot, P. *et al.* Inclusive fitness theory and eusociality. *Nature* **471,** E1–E4 (2011).
13. Sumpter, D. J. T. The principles of collective animal behaviour. *Phil. Trans. R. Soc. Lond. B* **361,** 5–22 (2006).
14. Pruitt, J. N., Demes, K. W. & Dittrich-Reed, D. R. Temperature mediates shifts in individual aggressiveness, activity level, and social behavior in a spider. *Ethology* **117,** 318–325 (2011).
15. Watts, J. C., Ross, C. R. & Jones, T. C. Diel and life-history characteristics of personality: Consistency versus flexibility in relation to ecological change. *Anim. Behav.* **101,** 43–49 (2015).
16. Pruitt, J. N. Behavioural traits of colony founders affect the life history of their colonies. *Ecol. Lett.* **15,** 1026–1032 (2012).
17. Falconer, D. S. & McKay, T. F. *Introduction to Quantitative Genetics* (Prentice Hall, 1996).
18. Gordon, D. M. The rewards of restraint in the collective regulation of foraging by harvester ant colonies. *Nature* **498,** 91–93 (2013).
19. Gardner, A. & Grafen, A. Capturing the superorganism: A formal theory of group adaptation. *J. Evol. Biol.* **22,** 659–671 (2009).
20. Pruitt, J. N. & Riechert, S. E. How within-group behavioral variation and task efficiency enhance fitness in a social group. *Proc. R. Soc. Biol. B* **278,** 1209–1215 (2011).